|  |  |
| --- | --- |
|  | **2013** |
|  | Arizona Geological Survey |

|  |
| --- |
| **Sustainability plan for National Geothermal Data System (NGDS) operations** Prepared for boise state universityMay, 2013 |
|  |

Contents

[Executive Summary 3](#_Toc357086716)

[Introduction 7](#_Toc357086717)

[Background 7](#_Toc357086718)

[Existing structure and organization 9](#_Toc357086719)

[Evolving concepts and market 16](#_Toc357086720)

[Premises and assumptions 17](#_Toc357086721)

[Challenges in Sustainability 20](#_Toc357086722)

Providing Content: Technical, Data, Services, and Applications 20

[Operations 24](#_Toc357086724)

[Organization and Management 30](#_Toc357086725)

[Business Models and Plans 34](#_Toc357086726)

[Conclusions 38](#_Toc357086728)

[Recommended Solutions 38](#_Toc357086729)

[Implementation Plan 40](#_Toc357086730)

[Acknowledgements and Disclaimer 41](#_Toc357086731)

[References 42](#_Toc357086732)

[Attachments and Links 44](#_Toc357086733)

[Appendix I: Case Studies 46](#_Toc357086734)

[Domain Science Groups: CUAHSI, IRIS, Unidata, IGSN, OpenTopo 47](#_Toc357086735)

[IT Groups:](#_Toc357086736) [Apache Software, DataNet, DataONE, W3C, DPN, OpenDAP, IETF, DataBasin, DataCite, Open Knowledge Framework 63](#_Toc357086737)

[Federated Groups: ESIP, OGC 81](#_Toc357086738)

[Appendix II: USGIN Protocols, Specifications, and Community Standards 85](#_Toc357086740)

[Category: Applications 85](#_Toc357086741)

[Appendix III: Recommendations for the Future of the U.S. Geoscience Information Network 87](#_Toc357086742)

[Introduction 87](#_Toc357086743)

[Background 87](#_Toc357086745)

[USGIN Today 89](#_Toc357086746)

[Achieving Sustainability: A Five-Year Vision 91](#_Toc357086747)

[Implementation 93](#_Toc357086748)

[Establish a Long-Term Governance Model 93](#_Toc357086749)

[Develop a Business Model 93](#_Toc357086750)

[Explore Testbed Opportunities 94](#_Toc357086751)

[Develop Marketing Strategy 95](#_Toc357086752)

[Costs 95](#_Toc357086753)

[Summary 96](#_Toc357086754)

[References 97](#_Toc357086755)

# Executive Summary

Since the 2009 American Recovery and Reinvestment Act (ARRA) the U.S. Department of Energy’s Geothermal Technologies Office has funded $33.7 million for multiple data digitization and aggregation projects focused on making vast amounts of geothermal relevant data available to industry for advancing geothermal exploration. These projects are collectively part of the National Geothermal Data System (NGDS), a distributed, networked system for maintaining, sharing, and accessing data in an effort to lower the levelized cost of electricity (LCOE). Determining “who owns” and “who maintains” the NGDS and its data nodes (repositories in the distributed system) is yet to be determined. However, the investment in building and populating the NGDS has been substantial, both in terms of dollars and time; it is critical that this investment be protected by ensuring sustainability of the data, the software and systems, and the accessibility of the data. Only then, will the benefits be fully realized.

The NGDS vision is such that the functions, accessibility, and availability of data will encourage continued participation within the NGDS, from all end users – data consumers, data providers, and application developers. As each node adds to is data repositories, the NGDS functions become increasingly valuable to it. Each data provider will have created a value added service that is transportable and scalable to cover all data in its possession. Thus, there are benefits to each participant to continue to add data to the system and maintain it. In the long term, we expect the data network to reach a ‘tipping point’ at which it becomes like a data equivalent to the World Wide Web – where everyone will maintain the function because it is expected and it fulfills critical needs. Applying this vision to the NGDS, it also opens the door for additional data providers external to geothermal development, thus wholly increasing the value of NGDS and its underlying data integration platform USGIN.

It is anticipated that NGDS will not be in full independent sustainable mode immediately at the end of the DOE ARRA funding in May, 2014, however. This is due to a number of factors including ambiguity about the ownership and functionality of NGDS prior to the redirection of the Design-Build project. We expect there will be a transition period from DOE sponsorship to full sustainability. The period and scope of that transition is dependent on the decisions made in the sustainability planning process and success in achieving them.

**Summary Conclusions**

The future of NGDS depends on taking measured steps to build this initiative and see it succeed. A summary of recommended actions from this document include:

* Develop and deploy a marketing strategy to increase engagement of the State Surveys, other data and service providers, and stakeholders to cement a strong user base supporting NGDS (and USGIN).
* Establish a governance structure led by a working group with representatives from NGDS participants and stakeholders.
* Decide if NGDS should be established as an independent non-profit 501(c)(3) foundation or a subdivision of an existing organization
* Conduct a comparative analysis of preferred sustainability models and select one to implement that will provide a conduit for financial resources to support infrastructure, a small (2-3 person?) staff, and servers for system infrastructure
* Inventory the State and Federal Geological Survey information architecture landscape to identify and prioritize use cases, and to identify capabilities and expertise to develop prototypes and testbed opportunities.
* Work to promote consistent best practices.
* Engage with domain science and computer science communities to promote interests of NGDS/USGIN in national cyber-infrastructure development.
* Keeping this operational system sustainable will require four core elements: continued serving of data and applications (technical sustainability); maintenance of system operations (operational sustainability); a governance structure (organizational sustainability); and an effective business model (business sustainability. Each of these presents a number of challenges.

**Recommended Solutions**

We propose preparing a formal business model based on NGDS being licensed by DOE to a newly formed non-profit USGIN for some period to get them both established financially and functionally. The business model should be vetted with the broad geothermal community and other potential sponsors, supporters, and clients of USGIN and NGDS services and data.

Client of USGIN

We propose that NGDS and USGIN be jointly developed at least during the initial phase of making

NGDS sustainable. NGDS is developed on the USGIN framework which is still emerging and maturing. NGDS serves as the largest client of USGIN services and is driving much of the USGIN technical development and data provider network (i.e., through AASG). Until both USGIN and NGDS reach full operational status, it is prudent to have them closely tied. Within a few years when the systems are each more of an operational commodity we anticipate that NGDS could be easily decoupled from USGIN and run independently, if there was reason or desire to do that.

USGIN Non-profit

The rationale and justifications for taking USGIN to non-profit status (Hutchison and Richard, 2011) are equally applicable to NGDS. This structure opens multiple opportunities for adoption and support by government agencies, funding from agencies and foundations, and even support from industry. Non-profit status does not prevent the setting up of a for-profit spin-off at a later time for commercialization of products or services if warranted.

**Broad-based stakeholder management**

Engagement of other end users and data contributors of the system will be facilitated by providing those entities opportunities for meaningful participation in the governance of the system. This may be through membership, advisory committees, boards, or comparable mechanisms.

Re-brand as Geothermal Data System (GDS)

NGDS is funded by US DOE GTO to catalyze and facilitate geothermal energy exploration and development in the U.S. However, there is a growing international interest in adopting not only the NGDS technical approach but NGDS itself. The Canadian Geothermal Energy Association has test demonstration data set served through NGDS and is reviewing what is required to establish a full service node on NGDS serving Canadian data.

The U.S. State Department proposed that NGDS be deployed to Latin America and East Africa, the latter in cooperation with the recently established East African Geothermal Program financed by USAID through the US Energy Association. Lawrence Berkeley Lab is developing a 10 week long geothermal training course for the EAGP to begin in Summer, 2013 and asked NGDS to provide a week long segment on data management. As NGDS provides more data and services and becomes more widely known and used, we anticipate that other adopters will similarly come forward. Renaming it to the “Geothermal Data System” would demonstrate the global capabilities of the system, reduce perceptions that it is only for US providers and users, thus broadening the user base and potential supporters. Individual nodes (e.g., countries) could brand their deployments of the GDS as [country name]GDS such as “Canadian Geothermal Data System (CGDS)” within the larger global network, effectively as a franchise.

Diversified client base

The NGDS targeted clientele is the U.S. geothermal industry and most of the NGDS data and services focus on the exploration and development parts of the industry. While that is a substantial economic entity it is still tiny compared to the potential natural resources, geoscience, and geospatial customer bases for NGDS data and services.

Diversified funding streams

We believe that a multifaceted approach to funding NGDS will be required. Not all the funding mechanisms we list will be viable, and in fact, some are contradictory or exclusionary. Nor will all of the viable mechanisms need to be or can be implemented initially. Our situation is not dissimilar to that of many Web-based businesses. They use a variety of business models, they experiment with them, and add or drop revenue streams as part of a highly entrepreneurial and evolutionary process. We need to emulate that same flexibility and innovation.

**Implementation Plan**

Decision Process

The decisions on how to proceed will be made by US DOE GTO, primarily in consultation with BSU and AZGS, the latter acting on behalf of AASG. If affiliation with USGIN is desired, AASG and USGS need to be involved.

Responsibilities

Creation of a non-profit USGIN will be led by AZGS for AASG. It is the intent of AASG and USGS for this to happen, regardless of what DOE decides on NGDS. If it is agreed that NGDS is owned by DOE, then they will need to determine how to delegate management of it via a contract, license, etc.

Alternatively, if DOE determines that NGDS is a product of BSU or BSU and the other NGDS project organizations, those entities will be solely responsible for decision making. However, given DOE’s substantial financial and intellectual investment in NGDS and anticipated compelling interest in seeing it succeed, it is expected that DOE will continue to have significant input (and investment) into operations.

Timeline

Decisions on the four areas of sustainable need to be finalized as quickly as possible in order to start the processes for technical, organizational, operational, and financial sustainability. The concern is that we cannot achieve these by the time the DOE funded projects end in spring 2014. There will be some transition period after April 2014 but we cannot predict the length and scope of that until the sustainability plan is adopted.

Resources

The AASG project has set aside about $250,000 in funding that was returned from project subcontractors who mostly could not complete the original scopes of work. We propose allocating that to implementing the business model for NGDS. This includes creation of informational, promotional, and educational materials, and hiring of a development manager, possibly based in Washington DC area, to line up additional end-users and system participants. Care will be taken to ensure that no federal funds are used in seeking additional federal financial support.

# Introduction

### Background

The creation of the National Geothermal Database began following a 2008, Deloitte Development, LLC, report which proposed that mitigation strategies to overcome barriers to development and enable additional investment in conventional and Enhanced Geothermal Systems (EGS) were needed to increase the use of geothermal as a base-load energy (report published for the U.S. Department of Energy (DOE) Geothermal Technologies Program (now Office, GTO) entitled *Geothermal Risk Management Strategies Report*). Specifically point 7.2 called for a National Geothermal Database to improve the accuracy, reliability, and general availability of geothermal relevant information (Deloitte, 2008). In addition, the report cited a 2000 National Renewable Energy Lab (NREL) study indicating that over a 25-year period, numerous geothermal research efforts were conducted with state and federal funding; however accessing that data was difficult and required significant resources. As a result, and in response to the risk mitigation strategies, inNovember of 2008, the U.S. Department of Energy released the Funding Opportunity Announcement (FOA) for the “National Geothermal Data Base” that will “serve as a central repository for all publically accessible geothermal data” (DOE, 2008).

The National Geothermal Database was charged with storing critical geothermal site attribute information such as temperature at depth, seismicity/microseismicity, fracture maps, drilling data, permeability data, well logs, geophysical surveys, and more. The database was to be inclusive of all types of geothermal resources including hydrothermal, geopressured, EGS, and coproduction with oil and/or gas. In addition, the database should utilize the existing USGS geothermal resource assessments and DOE funded research and development projects (DOE, 2008). The database was to be publicly accessible and serve to focus geothermal exploration activities, thereby mitigating investment risk. To maximize effectiveness, historical drilling information from the private sector was to be included.

*“DOE envisions creating an expansive, useful, user-friendly database   
that will continue to operate beyond the project period,   
even after all DOE funds have been expended” (DOE, 2008).*

Additional criteria from the FOA included the desire for flexible architecture so that data can be retrieved by a variety of factors and organized using a set of common metrics for assessing and comparing resources. This includes standard data collection units and methodology for data evaluation and documentation of the origin of the data to verify data quality, including metadata pertaining to the data set or collection of data items. An interactive mapping capability was preferred as part of the database to maximize utility and provide for user-friendly data visualization for three main areas – geothermal resources, institutional barriers (i.e. transmission lines), and power plant data.

According to the Deloitte report, “information and knowledge increase resource certainty and reduce risk, which allows for better access to capital, as well as more accessible and manageable base from which to develop a portfolio of projects or manage speculation across several leases.” A well-documented database will improve access to capital and potential investors due to the fact that it will provided additional resource certainty. Deloitte concludes that the database, “would be significant for conventional geothermal by reducing both investor uncertainty and upfront risk. It would be absolutely critical for the future EGS for the same reasons in addition to creating a foundation for a knowledge management system and success in the future” (Deloitte, 2008). Total estimate for the National Geothermal Database was $5 million to $7 million, per the Deloitte report, and the FOA dedicated $5 million to the effort.

**National Geothermal Database [System] Proposal**

In response to the 2008 FOA a collaborative group, known as the “Geothermal Data Consortium” responded with a proposal to build a system by using and adapting existing technology in coordination with emerging informatics standards and protocols while maximizing data availability by partnering with a series of academic geothermal centers to create a “National Geoinformatics System.” The project participants included: Boise State University, University of Utah’s Energy & Geoscience Institute, the Oregon Institute of Technology’s Geo-Heat Center, the Great Basin Center for Geothermal Energy at the University of Nevada, Reno, Stanford University, and the Association of American State Geologist’s (AASG) U.S. Geoscience Information Network (USGIN) as the system architecture.

The premise of the USGIN powered NGDS is that the system is open-source, web-based, and distributed. The distributed nature of the system is intended to support the long-term viability of the system beyond DOE funding as it emphasizes data stewardship and ownership at the institution where the data is housed. Current stakeholders and project participants already put considerable resources and personnel effort into their data sets and data sites; thus, by permitting the data to maintain at the data originators site, there is added incentive to maintain that data in the system. Thus, the proposal call for a “data system” rather than a “database” indicating that “the phrase ‘data system’ is used to reflect the need to build the next generation digital information system that is based on new and emerging technologies, standards and protocols…we must provide the tools and environment that truly give users a friendly and more powerful interface to a comprehensive data set” (GDC, 2009). One of the key components of the proposal was to provide for a self-sustaining mechanism that encourages people, groups, organizations, and industry, to put their information into the system as well as retrieve information, or a two-way system of “data-input” and “data-output.”

Included in the input/output design of the system was the inclusion of developers and programmers as a critical stakeholder. Open source projects, such as the Apache Software Foundation and Mozilla Foundation, rely heavily on code banks and forums where individuals interested in improving upon the software can participate. In fact, the Apache Software Foundation’s (ASF) motto is “community-led development since 1999 (<http://www.apache.org>) and the Mozilla Foundation promotes, “openness, innovation, and participation on the Internet” (<http://www.mozilla.org/foundation>). Further analysis of organizations like ASF and Mozilla Foundation are included in Appendix I. In order to meet the end-results of each stakeholder three generalized communities were identified, including: 1) those that use the data (e.g. industry, researchers, agencies, legislators, educators), 2) those that contribute data (e.g. researchers, industry, agencies), and 3) those that want to create applications (client side development) and tools for working with the data (e.g. software developers).

### Existing structure and organization

**Management & Contributors**

Shortly after the initial FOA was released, upon urging from President Obama, Congress passed the American Recovery and Reinvestment Act of 2009 (ARRA, February of 2009). ARRA significantly increased the available funding for renewable energy research, including geothermal. Since ARRA’s passage, DOE GTO has funded more than $33 million for multiple data digitization and aggregation projects under the NGDS through five projects which house multiple data “nodes” or data aggregators housing data for the system.

There are currently five projects, directly associated with the NGDS, one of which (the Geothermal Data Repository) serves as a data aggregator for all DOE GTO funded research projects. The five projects include three cooperative agreements and two interagency/interdepartmental agreements including:

1. **National Geothermal Data System Architecture, Design, Testing, and Maintenance**, PI Institution: Boise State University (BSU), Synopsis: responsible for the systems architecture and data compilation from the University of Nevada Reno, Oregon Institute of Technology’s Geo-Heat Center, University of Utah’s Energy and Geoscience Institute (EGI), and Stanford University. EGI anticipates acting as a node prior to the end of this project, University of Nevada, Reno is currently a node under the AASG project (#2 below)
2. **Association of American State Geologists, State Geological Survey Contributions to the NGDS**, PI Institution: Arizona Geological Survey (AZGS), Synopsis: responsible for the data digitization and integration from the fifty state geological surveys in addition to a small amount of new data collection. Nodes include multiple state surveys that are currently serving their own data in web accessible, machine-readable formats, and four data hubs deployed by region to accept data from the state surveys unable to host their own services. The hubs are located at AZGS, Illinois Geological Survey, Kentucky Geological Survey, and the Nevada Bureau of Mines and Geology. All but AZGS are located on the campus of a state university.
3. **Heat Flow Database Expansion for NGDS Data Development, Collection, and Maintenance**, PI Institution: Southern Methodist University (SMU), Synopsis: responsible for updating the SMU Heat Flow database with contributions from the Siemens Corporate Technology (SCT), Bureau of Economic Geology at the University of Texas at Austin, Cornell Energy Institute at Cornell University, Geothermal Resources Council, MLKay Technologies, Texas Tech University, and University of North Dakota. All institutions are providing data with the exception of SCT which is providing database and data analysis work.
4. **Geothermal Resource Classification**, PI Institution: USGS, Synopsis: an update to the 2008 Geothermal Resource Classification for inclusion in the NGDS.
5. **Geothermal Data Repository**, PI Institution: National Renewable Energy Laboratory, Synopsis: responsible for curating and hosting GTO funded projects data for submission to the NGDS.

The distributed nature of the funding and data population of this project is similar to the system itself.

**Systems Architecture**

The NGDS is designed based on a service-oriented approach using open standards to support data access by a wide variety of software applications, promote novel approaches to data analysis, and foster the development of tools by third parties. NGDS is based on the U.S. Geoscience Information Network (USGIN) which enables users to efficiently find, access, and share geoscience data, reducing the time and effort spent locating and integrating useful information and document new data by providing for information registration (by providers) and discovery (by users) based on standardized catalog services and metadata. The system accommodates resources in various forms, from unstructured text and images to documented, community Web services and interchange formats. To simplify, USGIN is a collection of Web-accessible resources that are registered in online catalogs and conform to data-sharing practices. Because USGIN is a network, it is largely intangible and rather defined by the protocols, interchange formats, and conventions that enable its operation. Thus the documents that define these protocols, interchange formats, and conventions ar the artifacts that represent the network. These documents are archived on the USGIN repository (<http://repository.usgin.org>), indexed through the USGIN Catalog, and accessed using standard Web resource retrieval. Links to the major specifications documents are also provided in the references section of this document.

The core components of the network are information exchange specifications and the catalog function. Information exchange specifications are community agreements on the conventions necessary for the interoperable exchange of some particular information. The exchange specification specifies a scope, the content model (descriptive components) for the data items of interest, interchange formats for encoding and transmitting information electronically, and the protocols used to request information. By using an interchange exchange specification a data provider can publish data that will be available to any client that implements the exchange, and a client application can access data from any provider publishing data according to the exchange conventions. This allows data in a variety of formats, organization, and structure to be integrated without having to manually transform it.

The catalog function comprises a collection of metadata records that describe resources accessible through the network (in this case, USGIN), and a special information exchange that defines metadata content, how the metadata collection is searched, and how metadata are encoded in search responses.

As a system, USGIN is a loosely coupled system of independent data providers, client applications, and infrastructure. The infrastructure includes 1) tools for registration of new resources, searching metadata catalogs, authentication, and resource validation; 2) registries for vocabularies, agents, specifications, and interchange schema; and 3) documentation and educational resources. Because network operation is based on information exchange specifications that are independent of any particular hardware or software implementation, all of the operational components can evolve as technology evolves. Use of standard protocols enables data access using off-the-shelf software, both commercial and open-source. Open Geospatial Consortium (OGC) Web Map Service (WMS) and Web Feature Service (WFS), and OpenDAP NetCDF services are specified in current USGIN information exchanges. These services were specifically chosen due to the open-source nature and existing developer community. Details on the use of these services, and interchange formats for data exchange, are can be found in the References & Links section of this document.

The distributed nature of the system means that stewardship of resources is determined by the resource owner. Participation in the network only requires that a resource provider create metadata that conforms to the USGIN profile, and make the metadata and the described resource available. The network is open; anyone can deploy new nodes and components that implement one or more USGIN specifications, without requiring approval. New specifications can be introduced for service protocols, interchange formats, or vocabularies. Keeping resources under the stewardship of the parties responsible for the information promotes system sustainability because the stewards have a direct connection with the quality of the product rather than submitting the data to a centralized database and related data manager.

***Elements of the Network:***

* ***Catalog***: the catalog is a collection of metadata records describing resources that are intended to be considered part of USGIN. A catalog should be thought of as a cloud of records, there may be no individual data-store on the network that contains the entire collection of metadata as catalogs can harvest from one another (improving functionality and search capability) and expose the data to commercial search engines (e.g. through OGC’s Catalog Service for the Web, CSW). The use of a standardized encoding scheme for metadata interchange and a standard service protocol for accessing catalog content simplifies interoperability and enables the use of off-the-shelf software to implement the catalog system. The scope of the catalog includes data products or datasets; these are units of information that have been authored, edited or compiled under some stewardship, with some purpose and procedure, and having a common collection of access processes.
* ***Metadata:*** the data about the data. Individual documents require one metadata record per document. These documents might be scans of well logs, scanned reports, maps, or publications, multimedia files, or data compilation spreadsheets (e.g. Excel or Access file). Some document types may consist of a bundle of files, e.g. Esri shape file. In general, these should be bundled into a single file like a zip archive or UNIX tar file. The metadata must include the URL at which the document can be accessed, if it is not accessible online, the metadata should provide instructions on how to access the document. Full metadata documentation can be found in the linked resources section of this document under USGIN Standards and Protocols Drafting Team, 2010. In order to maximize usability and interoperability, we have adopted ISO metadata schemas for geographic information, ISO 19115 and 19116 as well as ISO 19139 for XML encoding.
* ***Information Exchanges***: an information exchange is an implied contract that a data provider follows to expose a particular kind of data through a particular service protocol in a particular interchange format. Each information exchange specification defines one or more Web services. The development of information exchange specifications for USGIN are based on the operational systems of OGC, ISO, the Environmental Protection Agency’s Environmental Information Network, and the NGDS. Content models are abstract specifications that define a feature for information interchange, along with the properties associated with the feature and relationship to other features. For example, a content model featuring Bottom Hole Temperatures for boreholes will provide the temperature of the borehole, depth at which the temperature was observed, and whether the temperature is in Celsius or Fahrenheit (along with other critical features of the borehole, such as location). When data providers follow this format, it ensures that complete data for the feature is available. Once data is portrayed in a content model, it must be implemented using an encoding scheme that is computer processable. The eXtensible Markup Language (XML) is typically used to implement the content models, but other schemes (.e.g JSON) may be adopted as technology evolves. The important thing is that the information is encoded consistently in a structured, well-documented format. This enables computer programs to parse the interchange documents and extract the desired information, and as newer encodings are adopted, conversion between formats can be automated using software. These information exchanges help implement truly interoperable data (additional information discussed below).
* ***Community Specifications***: One of the operating principles of USGIN is to not “reinvent the wheel,” i.e. to use and extend existing software whenever possible instead of developing new software. Active, free, open-source software communities offer a path to long term viability of key system components whether or not individual projects continue. USGIN specifications, in general, define conventions and practices for the use of existing components and standards to simplify interoperability.

Appendix II provides a table of community standards, profiles, protocols, and software employed by USGIN, and subsequently NGDS, to address the challenges associated with sharing large amounts of diverse data stored in geographically disparate locations. In general, these are not the only solutions available, but indicate the resources that the systems architects have found useful.

The NGDS combines compatible client and server side applications to create a nationwide distributed network of geothermal relevant data providers. Currently, data providers are implementing a variety of client and server applications for data management. This means that nearly anyone’s data management system is compatible with the NGDS, and linking the most basic data to NGDS would require metadata registered in an NGDS compliant catalog. Recognizing, though, that not all individuals interested in participating and serving data into the system are able to or interested in selecting individual server side implementations, NGDS developers are creating a simplified software stack for server side implementation, a “Node-in-a-Box” (NIAB).

**A Simplified Software Stack, Node-in-a-Box**

The NIAB software targets organizations or individuals who wish to host at least one of the following:

* An online repository containing resources for the NGDS;
* An online site for creating metadata to register resources with the NGDS
* NGDS-conformant Web APIs (application programming interface) that enable access to NGDS data (WMS, WFS, WCS)
* NGDS-Conformant Web APIs that support the discovery of NGDS resources via catalog service (CSW)
* A web site that supports discovery and understanding of NGDS resources.

A number of different open-source frameworks for development of this online application were reviewed including the existing systems in use for the AASG project and SMU project, as well as CKAN, GeoNode, and DSpace. However, the NGDS Design and Testing Team determined to use CKAN (<http://ckan.org>), developed by the Open Knowledge Foundation, because it provides the closest match between out-of-the-box functionality and NGDS NIAB requirements (Clark et. al, 2013; NGDS Systems Architecture, 2012). Papers documenting the CKAN implementation were presented at the 38th Annual Stanford Geothermal Workshop and are in preparation for the 2013 Geothermal Resources Council Annual Meeting.

A major consideration in the evaluation of the software system was the long-term viability (sustainability) of the eventual application. By building on an existing, active and widely used open-source project, it ensures that external developers can assist in the ongoing maintenance of the system. Other factors considered in the development framework included; 1) adaptability of the user interface to be compatible with an independently developed user experience concept for NGDS users; 2) ease of extensibility, with a plug-in architecture that allows addition of functionality without having to modify the core codebase; 3) support for geographic data and map-based search and data browsing; 4) support for administrative activities like user management, access control, and activity logging (Clark, et.al, 2013).

CKAN is written in Python and makes use of a variety of open source frameworks including Pylons, which itself is a combination of various open source frameworks integrated to form the basis for Web-based Enterprise-level applications (Clark, et. al., 2013). The primary CKAN user scenario is data storage and management and includes file storage, metadata management, and management of structured data. In addition, it offers a plug-in mechanism enabling developers to rapidly extend CKAN’s core functionality; most importantly to the NGDS, the ability to support geographic features as well as exposing metadata according to the OGC standard catalog (CSW) (Clark, et. al., 2013). CKAN implements crucial housekeeping features such as user management and logging, which can be tedious to implement but crucial for the site’s uasability. Finally, CKAN has been selected as a candidate for the next implementation of Data.gov, the site designated to help increase the availability of the public to easily find, download, and use datasets that are generated and held by the Federal Government ([www.data.gov](http://www.data.gov)); thus, adding to the assumption that a CKAN based software stack will encourage long-term use by the three core NGDS stakeholders: data users, data contributors, and developers.

While there are many attributes to the out-of-the-box CKAN functionality, there are a number of items that required NGDS developer modification including (Clark, et. al., 2013):

1. Modification of the User Interface to reflect the user experience testing as completed through the NGDS Architecture, Design & Testing project
2. Creation of the NGDS/USGIN metadata, including spatial extent
3. Consistency check for well-known structured data files (described in the next section)
4. Providing OGC services for uploaded structure files with geospatial information
5. Full-text indexing of documents
6. Role-based right for uploading and publishing data
7. User feedback and rating of uploaded data

Once completed, the code will be registered and stored on the USGIN GitHub code sharing repository (<https://github.com/usgin>) as well as registered with the CKAN developer code repository (<https://github.com/okfn/ckan>).

**Data**

NGDS data consists of structured (e.g. tabulated and formatted data representing features or observations like well headers, heat flow, borehole temperature, etc.) and unstructured (e.g. single file format such as PDF publications and other documents). As location (geographic) data is critical to the system (i.e. evaluating site specific areas for geothermal exploration), most of the structured data is represented in geographic information systems (GIS) data through OGC Web Map Services (WMS), and Web Feature Services (WFS). Unstructured data is location bound through one of the metadata entries to assist in location based search functionality.

To aide in interoperability, a series of geothermal relevant information exchange templates were developed. These were developed with input from the geothermal end-user community by reviewing and implementing existing formats whenever possible. The exchanges were reviewed through the Geothermal Data System Design and Population Working Group (GDSDPWG) prior to implementation on the NGDS site, [www.geothermaldata.org](http://www.geothermaldata.org). The information exchanges are also available via (<http://schemas.usgin.org/home/>) or the USGIN GitHub.

Current data exchange models include:

* Active Fault/Quaternary Fault
* Aqueous Chemistry
* Borehole Lithology Intercepts
* Borehole Lithology Interval Feature
* Borehole Temperature Observation
* Direct Use Feature
* Drill Stem Test Observations
* Drill Stem Test Observations (deprecated as it is now combined with the Well Observations)
* Fault Feature/Shear Displacement Structure
* Fluid Flux Injection and Disposal
* Geologic Contact Feature
* Geologic Unites
* Geothermal Area
* Geothermal Fluid Production
* Geothermal Metadata Compilation
* Geothermal Power Plant Facility
* Heat Flow
* Heat Pump Facility
* Physical Sample
* Powell & Cummings Geothermometry
* Power Plan Production
* Radiogenic Heat Production
* Rock Chemistry
* Seismic Event Hypocenter
* Thermal Conductivity Observation
* Thermal/Hot Spring Feature
* Volcanic Vents
* Well Fluid Production
* Well Header Observation
* Well Log Observation
* Well TestsData exchange models maximize interoperability and while heavily encouraged by the development team, they are not required to participate in the NGDS. NGDS data can take the form of three distinct tiers of data interoperability, including:
* Tier One – unstructured data, with metadata describing the location and type of information contained in the file,
* Tier Two – structured data, with metadata describing the location and type of information contained in the file,
* Tier Three – structured data in a standardized defined format, with metadata describing the location and type of information contained in the file.

Table 1 provides additional information on the input/output requirements of each tier.

|  |  |  |  |
| --- | --- | --- | --- |
| **Tier** | **Description & Example** | **Input Requirements** | **Output Requirements** |
| 1 | Unstructured Data, for example scanned documents. E.g. PDF scan of a scientific report with tables of heat flow data in the appendix. | Upload the file to a web accessible location; Create metadata for record. | Accessible via HTTP GET; Metadata points to the URL accessible via CSW |
| 2 | Structured (tabular) data not conforming to an NGDS interchange format. Examples include Excel/csv tables, Access, etc. E.g. a MS Excel spreadsheet containing the heat flow data in a format defined by the researcher collecting the data. | Upload files to a web accessible location; Create metadata for the record. Metadata should include descriptive information about the column headers and data types that the documents contain. | Accessible via HTTP GET; Metadata points to them via CSW. Optional – data may be exposed as an WMS/WFS. |
| 3 | Structured data conforming to an NGDS content model. E.g. a CSV file or Excel file that can be transformed to XML for machine consumption in a specified format. | Upload files to a web accessible location; WFS and WMS services are deployed to serve data; Create metadata pointing at all endpoints (because the content model defines entity/attribute information the creation of metadata is less cumbersome). | Accessible via HTTP GET; Data exposed as WFS/WMS services; Metadata points to all endpoints accessible via CSW. |

Table . Input/output requirements in Each Tier

### Evolving concepts and market

Subsequent to the Deloitte Report, additional recommendations from both national and international experts have called for the development of a shared knowledge management system for geothermal data. In March of 2011, DOE convened a panel of geothermal experts in Albuquerque, NM for a guided discussion on the future of geothermal energy in the U.S (U.S. DOE, Office of Energy Efficiency and Renewable Energy, 2011). One of the recommendations from the 15 panelists was that DOE focus its resources on reducing the cost of confirming known hydrothermal resources and identifying undiscovered hydrothermal resources. In order to do this, the panel suggested developing an inventory of prospects using existing technology. Also in 2011, the International Energy Agency released the *Technology Roadmap: Geothermal Heat and Power* which called for “develop[ing] publicly available databases, protocols, and tools for geothermal resource assessment and ongoing reservoir management to help spread expertise and accelerate development” (International Energy Agency, 2011). When completed, NGDS will help achieve both reports’ recommendations.

The concepts of “Big Data” (large, complex data sets) and “Open Data” (certain data sets should be freely available to everyone to use and republish) have compounded recently as nations across the world are interested in sharing publicly funded data to improve and enhance public sector innovation. NGDS is ahead of these initiatives in that it provides an answer to both.

Recently, three documents on open data policies at the federal level were released. These include an Executive Order, and memos from the Office of Science and Technology Policy (OSTP) and Office of Management and Budget (OMB).

The OSTP Memorandum for Heads of Executive Departments and Agencies, *Increasing Access to the Results of Federally Funded Scientific Research* in February 2013 ensures that federally funded research is “available to and useful for the public, industry, and scientific community” (Holdren, 2013). This requires agencies with research budgets over $100 million to ensure that “…the direct results of federally funded scientific research are made available to and useful for the public, industry, and the scientific community. Such results include peer-reviewed publications and digital data” (Holdren, 2013). The memo further requires that agencies ensure that publications and metadata produced with federal funds are stored in an archival solution that:

1. Provides for long-term preservation and access to the content without charge
2. Uses standards, widely available and, to the extent possible, nonproprietary archival formats for text and associated content (e.g. images, video, supporting data)

The Executive Order, released May 9, 2013, *Making Open and Machine Readable the New Default for Government Information*, calls for government information to be managed as an asset throughout its life cycle to “promote interoperability and openness, and, whenever possible and legally permissible, to ensure that data are released to the public in ways that make the data easy to find, accessible, and usable” (White House, 2013).

The OMB memo was also released on May 9, 2013 and begins the implementation of the Executive Order by requiring agencies to collect or create information in a way that “supports downstream information processing and dissemination activities” (OMB, 2013). This is specifically requested to assist and enable the development of products and services that benefit the public.

These reports indicate a continuing need for the geothermal community to have easy access to geothermally relevant data in order to reduce the costs of confirming hydrothermal and identifying EGS. The federal open data initiative promotes distributed data sharing and life cycle maintenance of federally funded datasets. Thus, the NGDS vision of shared and easy data access can assist DOE in completing these open-data requirements.

### Premises and assumptions

This plan assumes that the NGDS Architecture, Design, & Testing will finalize software development on the NIAB in September of 2013 with troubleshooting taking place in October of 2013. Minimal software development support will be available through April of 2014. Project sustainability and deployment of additional nodes and data collection will remain in full development through April of 2014. This will complete the DOE funded section of the NGDS. At that point, all projects currently funded by DOE will be linked through a catalog aggregator, which will be housed at AZGS.

**Existing Nodes**

SMU has authored a sustainability plan for the SMU heatflow database node (Blackwell, et. al., 2012), which indicates that the majority of the data will be housed on the SMU servers and through an agreement with SMU’s Office of Information Technology. AZGS data hubs are in the process of virtualizing their servers with Windows Server 2012 on Amazon Web Services; this is due to the replication and disaster discovery plan’s recommendation for Hyper-V server replication (AZGS, 2013). AZGS is encouraging the nodes to purchase the maximum three year term for Amazon Elastic Compute Cloud in order to maximize the availability of data while additional funding streams are identified and matured. EGI will be implementing the NIAB software stack.

**System Users (Allison, et. al., 2013, Blackman, et. al., 2013)**

The system is designed around three distinct users comprising a vast array of specialties and organizational structures. These include data providers, data consumers, and applications developers. We presume that the core stakeholders in NGDS will remain the same throughout the lifespan of the NGDS. However, as part of maintaining the NGDS, there may be members of related scientific communities interested in accessing or participating in such a data sharing environment (this premise will be discussed in detail in a later section).

**Data Consumers:** The primary end-users of the system. Those individuals and/or organizations interested in querying the information within the NGDS. These users require primarily “read-only” type access to the information on the NGDS and require primarily free accessibility to encourage use. Free access also limits the development/back-end technology required for maintaining user accounts, thus contributing to minimized costs of operation. A potential mechanism for funding the NGDS is the use of specialized “membership” or access to specially developed applications for a fee. This will be further discussed in the next section. While general use, system wide, will remain open and accessibly, individual nodes, or external data providers, may choose to implement logon-credentials for accessing their data. Again, discovery of the data should maintain free and accessible.

We foresee data consumers as members of the community from Industry, Researchers, Agencies, Legislators, Interested Public, Students & Educators, Financial & Investor Communities, among others.

The NGDS may need to incorporate a form of “crowd-source” rankings to assist data consumers in the identification of valid or high quality data. While NGDS is open for data submissions from multiple organizations and institutions, some data may have more value than others. For example, data collected by a private geothermal exploration firm may be more valuable than data collected from an undergraduate/citizen science experiment. However, if the only data available for a location is the citizen science experiment, NGDS should provide access to that data.

**Data Providers:** Individuals and/or organizations with data (Tier 1, 2 or 3) to load onto a node of the NGDS for access by general users OR individuals/organizations who would like to contribute data by becoming a node.

Users who wish to host their own node on the NGDS could potentially require some assistance by a centralized group rather than solely relying on developer documentation available at the NGDS Web site and code sharing hub. There may also be the requirement of a long-term ‘broker’ or individual that can assist with data transformation. Currently this is being handled by the individual nodes. The vision, though, is that data providers will recognize the value of portraying their data in Tier 3 format for interoperability. This is not an unlikely vision as the World Wide Web has similar standards and protocols for ‘participating’ in the web, this is HTML.

Some nodes may have the need to distinguish among roles within this category, such as those who may add to data collections from those who may remove or replace data. In any event, administering this type of user community is likely to require knowledge of the individual(s) contributing data, where they are located, under whose authority they are acting, etc. Maintaining this function will be ongoing and will likely require helpdesk type support capabilities or an ‘node administrator.’ This could range from resetting passwords, to providing guidance on file uploading procedures, sources of documentation for software questions, etc. The repository node through which external contributors upload information will likely need to be prepared to support this function.

**Application Developers:** Individuals or organizations interested in developing additional functionality to the NGDS and maintaining its open source code. These users are likely to have more information technology experience than either the data consumer or the data provider. While it would not be necessary to know where these users were located or under whose authority they are acting, they will likely have questions about the data format layouts and communication protocols used. A variety of helpdesk type support might be necessary, but could be reduced by well written easily accessible documentation, or through the development of online discussion forums as are typical in developer environments, such as GitHub repositories.

At the time of this writing, we assume that upon April 30, 2014 all DOE funding for NGDS will end and at that point, NGDS must be self-reliant.

# Challenges in Sustainability

In order to better focus efforts to promote long-term sustainability of NGDS, we first explore four challenges in achieving sustainability: providing content, operations, organization and management, and business models and plans.

## Providing Content: Technical, data, services, & applications

The NGDS vision is such that the functions, accessibility, and availability of data will encourage continued participation within the NGDS, from all end users – data consumers, data providers, and application developers. Once the protocols and standards are in place in each data node, additional data added to each database are automatically promulgated to the system. As each node adds to is data repositories, the NGDS functions become increasingly valuable to it. Each data provider will have created a value added service that is transportable and scalable to cover all data in its possession. Thus, there are benefits to each participant to continue to add data to the system and maintain it. In the long term, we expect the data network to reach a ‘tipping point’ at which it becomes like a data equivalent to the World Wide Web – where everyone will maintain the function because it is expected and it fulfills critical needs. Applying this vision to the NGDS, it also opens the door for additional data providers external to geothermal development, thus wholly increasing the value of NGDS and its underlying data integration platform USGIN.

In this sense, NGDS can be considered an “ecosystem” of users that evolve with time and development. The success of the ecosystem depends on the longevity of the software and all that relates to it, such as the data formats, protocols, etc. NGDS was specifically built on open-source and commonly used standards and protocols in order facilitate that opportunity for the software to evolve. Parties involved in the ecosystem can improve parts of the system by dedicating resources toward that end as fits their needs, especially if well documented. Crucial factors that enable this are:

* Usability and quality documentation which makes the system usable from the standpoint of the data provider and data consumer. Systems that are less easy to understand or otherwise present a barrier to entry without significant reward are less likely to flourish.
* Code quality and sound system documentation make the system usable from the standpoint of the software/application developer. Systems that are easy to build on are more likely to attract developers who always have the option of adopting other systems or building from scratch.

Often, at the onset, centrally guided efforts are necessary – such has been done with the NGDS – to develop the initial software and protocols. Long-term, however, an organization that aggregates the activities of long-term maintenance (without necessarily taking full responsibility for executing all activities) is very beneficial to the sustainability of the NGDS. As can be seen with other similar software ecosystems (e.g. The Apache Foundation, The Mozilla Foundation) they thrive based on the interest and involvement of actors in the ecosystem, but only by addressing the needs of the ecosystem.

A large portion of the NGDS work completed to date has been on the technical viability of the program – creating and implementing standards and protocols, building upon existing software systems, and encouraging data compilation for maximum interoperability.

**Content Sustainability**

Content sustainability is critical if users are to find the NGDS valuable for years to come. This is, the strength of the distributed network in that content is generally housed with the data provider, and thus maintained by that provider. The nodes of the NGDS must continue to be populated by current sources of raw data: public and private industry, governmental agencies, and educational institutions. Some measures that may help ensure NGDS content stays current include:

1. Continue to require all recipients of GTO funding are required to contribute certain types of data to the Geothermal Data Repository as a contingency of their federal support as GDR will be accessible via the NGDS.
2. Require or financially incent the next incarnation of data contributors by specifying the type of data contribution (Tier 2 and Tier 3, versus Tier 1 as described above)
3. Expand this model to other sources of public funding, including the National Science Foundation (NSF), and other programs areas of DOE such as the fossil fuel driven initiatives, USDA Rural Development grant recipients, etc.
4. Research and development that falls within Dr. Holdren’s Memo to heads of scientific research agencies may be encouraged to contribute data.

In order to maintain content viability, mechanisms for contributing revised datasets, or datasets relevant to a base (original) set is necessary. Thus, NGDS should provide for a mechanism for external, or “crowd-sourcing” of data to ensure maintenance. Linking data sets is currently on the development schedule for the NGDS Architecture, Design, and Testing team developers.

In the case of the SMU Node of the NGDS, the data contributors will have the ability to upload files in the same format as used previously to update or augment previous submissions. They will have specially designated user IDs that will enable them to see the last successfully uploaded file and/or upload new files. Once satisfied that the data is validated, they can elect to ‘publish’ it, thereby making it available to the rest of the NGDS and the public at large. Contributors to the SMU Node upload their files as ‘Tier 2’ structured data, which is then transformed by custom software written by Siemens Corporate Technology into one or more of the ‘Tier 3’ layouts specified for use in exchanging data via WFS on the NGDS.

Other nodes of the NGDS will handle this conversion differently, some of them through the use of installation of a NIAB. AASG nodes that are self-hosting can update their data at any point; although datasets housed in a repository or hub may require the assistance of staff – a “broker” – at one of the hubs. Depending on the outcome of some of the funding mechanisms identified in this plan, data from one or more of the hubs may be replicated and housed at an alternative location; that is, one or more of the AASG hubs may be shut down and their data storage transferred to hub location that is able to maintain the data. However, we expect that within the three year terms of service for the Amazon Web Services for housing and maintaining data, that we will be successful in identifying and cultivating additional revenues streams and thus, closing a hub should not be necessary.

**Providing Content**

By far, the largest amount of funding has been distributed to digitizing and providing interoperable content to the NGDS. All projects under the current NGDS program have content providers as part of the projects, the NGDS Architecture, Design, and Testing is the primary project focused on the systems operations and architecture.

Currently the incentives for participation are financial – more than 60 organizations are currently digitizing and distributing data to the NGDS through their nodes and hubs while registering metadata in the NGDS catalog. Generally, data maintenance is the responsibility of the data provider; however, in cases where the data provider is unable to maintain their data, separate hubs or repositories have been created to assist with data deployment and curation.

Additional incentives for participation occur when a “tipping point” of data transpires. That is, when the data available in the system becomes so valuable and easily accessible that it cannot be found elsewhere, thus users are inclined to continually access the NGDS for their research and exploration. We feel that this tipping point will occur when all currently funded data is contributed. End users must be aware of the system though in order for it to be functional; additional discussion on attracting end-users will occur in the Managerial Sustainability section of this report.

Thus, we envision that the NGDS will encourage continued participation once data contributors realize the impact of the combined resources. In addition, data contributors are encouraged to house and deploy their own Web services as members of the NGDS. As such, we are encouraging a paradigm shift in data management. Two decades ago organizations and agencies did not all have Web sites; today, it is a general cost of doing business and most organizations view a quality and easily accessible Web site as a general operational cost. We envision that with the tipping point in data aggregation, the Federal government’s Open Data initiative, and the movement to Web services in addition to Web sites, that the NGDS data will be maintained. Costs associated with data storage and maintenance are discussed in the Systems Operations component.

One mechanism that is currently in place for data contributions, that should continue, is the requirement that GTO funded projects contribute data to the Geothermal Data Repository and thus the NGDS. In order to clarify the data submission process, project officers should review the potential data that will be produced by the project and identify the best mechanisms for distribution, prior to releasing the award. This would result in a data management plan for geothermal projects funded by GTO. If at all possible, GTO should encourage and coordinate additional offices within EERE to evaluate additional data sets of utility for NGDS.

An additional option for providing content to the NGDS is through “data preservation” grants if a non-profit entity is established to house the NGDS. These could be small grants ($5,000 to $10,000) for preserving and maintaining distinct data sets. Elsevier Publishing and IEDA at Lamont-Doherty Earth Observatory, Columbia University, are currently collaborating to conduct similar small preservation awards. Collaborating with a scientific publisher is one mechanism for funding this type of mini-grant.

Providing content to the NGDS must be simple and relatively hassle-free. For this reason, we have implemented a relatively simple metadata schema for resource registration within the NGDS. Currently data providers can register resources through an independent node on the NGDS, by deploying their own node (through the specifications discussed in Appendix II or by deploying the NIAB software stack), or through independent XML validation and registration with the catalog. Maintaining the catalog and metadata registration is a requirement of NGDS sustainability.

Currently, the AASG repository has a simplified metadata creation tool available at <http://repository.stategeothermaldata.org> running on virtual machines through Amazon Web Services. A user-ID and logon authentication is required for accessing this repository. User-IDs are validated by an employee at the AZGS. A small “help-desk” style support system will be required for ongoing maintenance of the project; thus funding is required for an individual to maintain user-authentication and the repository. This can be completed with relatively minimal support, estimating approximately 4 weeks effort per year. Additional analysis will on costs of the node/repository maintenance are discussed in the Systems Operations section.

**Targeting Content**

In addition to the NGDS, DOE has commissioned the Geothermal Prospector, an application in development at NREL. Geothermal Prospector will be able to consume the WMS, WFS, WCS, from the NGDS catalog, as well as geolocation data for publications to assist with a national gap analysis of geothermal data. This assessment and tool will assist DOE in identifying additional funding locations and data types; thus adding to the ongoing data provision within NGDS.

NREL is also currently engaged in a Regulatory Roadmap for Geothermal Energy. Incorporating the findings of the NREL regulatory roadmap team with the NGDS data will assist geothermal exploration in identifying the individual federal, state, and local regulations with the exploration process. Combining regulatory factors with data analysis will also assist in risk mitigation, particularly for financial investors. Thus, efforts should be focused on ensuring that the Regulatory Roadmap is accessible for NGDS stakeholders. If this is an additional metadata entry required by the current NGDS development team, it should be identified prior to the Fall of 2013.

The following summarizes the key findings for Content Provision:

* Individual data services maintenance is part of a shift in data management similar to the creation of Web sites for organizations and business. Data should be valued as a similar, if not more important asset.
* Ensure that NGDS reaches a “tipping point” in terms of data availability and provide for end-user engagement and awareness of NGDS and its data attributes
* Continue the current GTO funding requirement of contributions to the NGDS through the GDR. Improve upon the process by requiring a data management plan to be determined between the project principal investigator and the DOE program officer prior to funding. Ensure that DOE program officers are aware of the different materials available for training funded project members (more fully discussed in the Systems Operations, Education, Outreach, & Training Section.
* Potential for ongoing financial incentives for data contributors through mini-grants if the final home for NGDS can accommodate this.
* Ensure that metadata registration is simple through sustaining at least one aggregator catalog node and support system.
* Utilize the Geothermal Prospector tool, currently in development at NREL, to identify community needs and data gaps.
* Link the NREL developed Regulatory Roadmap to the NGDS through quality metadata records to improve the risk mitigation for exploration as well as financial investors.

## Operations

**Systems Operations Longevity (Allison, et. al., 2013)**

NGDS is designed as an open-source, web-based, nationally distributed (vs centralized) system with diverse users and providers, making it scalable and transferable to data providers and users in other fields.

NGDS is built upon the data-sharing platform, U.S. Geoscience Information Network (USGIN) which is a series of open-source standards, protocols, and specifications funded by the National Science Foundation from 2008-2012. It is a collaboration between the U.S. Geological Survey and Association of American State Geologists resulting from a 2007 workshop recommending the creation of a distributed, national “Geological Information Network,” of digital Earth Science data using common standards and protocols, preserving ownership, credit, and control of data, and building on existing data systems (Allison & Dickinson, 2008). USGIN enables users to efficiently find, access, and share geoscience data by registering (and discovering) data using standardized catalog services and metadata. Because USGIN is a network, it is largely intangible and is rather defined by the protocols, interchange formats, and conventions that enable its operation (Richard, 2013). Documents developed under USGIN are now under review and maintained via a GitHub repository for public access. Using these documents, anyone can replicate, maintain, and update the data integration framework for NGDS.

The data system approach of individual web hosts and repositories, in conjunction with the networking capabilities that have developed and been implemented in NGDS are expected to be used subsequently for more and broader data than geothermal. This serves two functions critical to sustainability of the system. Because in addition to its basic goals and objectives, it adds permanent functions that will be used to develop the promises offered by the system architectural concept – internal and external interoperability and access to a growing number and array of third-party services, applications, and tools.

Once the protocols and standards are in place in each data node, additional data added to each data base are automatically promulgated to the system. As each state adds to its data bases, the NGDS functions become increasingly valuable to it. Each data provider will have created a value-added service that is transportable and scalable to cover all data in its possession. Thus, there are benefits to each participant to continue to add data to the system and maintain it. In the longer term, we expect the data network to reach a ‘tipping point’ at which it becomes like a data equivalent to the Web – everyone will maintain the function because it is expected and it fulfills critical needs.

The systems where the data resides must be maintainable. The use of the term ‘systems’ encompasses the hardware, operating system software, and network connectivity. This is complicated by the fact that there is no ‘one’ system or even one organization responsible for the ongoing maintenance of the various systems comprising the NGDS. AZGS is charged with the maintenance of the AASG server(s); DOE and NREL will maintain the Geothermal Data Repository; SMU will maintain their node (Geothermal Data Aggregation, GTDA), etc. While this may initially appear as a fault of the system, this is also the strength of the system.

Making use of open source systems and/or widely used commercial products will help provide interoperability, which is a necessary component for ongoing sustainability in a distributed system environment, but it does not fully address the myriad of issues that must be considered. In addition to selecting open source options where practical, utilizing ‘loosely-coupled’ components also aids sustainability. This means, that front-end, or client side applications, are not dependent upon an individual back-end, or server side application; thus, multiple client (or server) applications are possible. It is similar to the notion that a car owner should be able to replace their car battery without needing to replace the entire engine, and without needing to even understand how the engine works. To continue with the car analogy, just as there is an interstate highway system with standard ‘rules of the road’ understood by all types of drivers in all sorts of vehicles, even when passing through different state and local jurisdictions, the network that communicates among the NGDS nodes must be open to participation from a myriad of locations and platforms, provided they make use of certain protocols when communicating information.

**System Operations**

A weakness of the distributed network is that there are multiple contributors to the system and thus, multiple stakeholders to encourage continued participation. However, as this report has already discussed, it is an asset to the system as well, as it means that not all maintenance and development is assigned on a single point. Thus, it is a fail-safe mechanism in that if properly scoped and with proper replication procedures in place, the system should be able to absorb defunct data contributor’s assets.

AZGS is currently in the process of implementing a disaster recovery plan through data server replication. This is based on the premise of entirely virtualized server machines, primarily running on Amazon Web Services. While other options are available, we are suggesting the use of Microsoft Server 2012 and Hyper V replication. This permits virtual data back-ups and nearly instantaneous deployment if an issue arises.

Since NGDS is built using USGIN standards and protocols – a network built using open-source, web-based applications and supported by AASG and the USGS (through ScienceBase) – it invites open development through developer resources such as clearly documented and available code, discussion forums and blogs, and reference architecture documentation as well as white papers and conference proceedings on the system.

The USGIN sustainability study (Hutchison & Richard, 2011; Appendix III) identified what components are likely to be necessary to sustain USGIN:

Servers and software: (This list assumes that network infrastructure (internet connection, switches, firewalls, DNS, etc.) are in place)

* Catalog and repository server: Linux/Tomcat/PostgresQL/ GeoNetwork or Geoportal (could run on Windows stack as well). Only one is essential, any number is possible.
* To support a catalog system: a server for registries for identifiers and vocabularies, in addition to repositories for system specification documents and resources, like XML schemas that must be web accessible to support service operations.
* Data server: Windows/ArcGIS Server (dbms optional if shapefiles are used as data source), or Linux/Tomcat/PostGresQL/Geoserver or Mapserver. One server with modest capabilities could serve perhaps 300 Mb of data, depending on load.

Personnel:

* Technical IT personnel: Need capabilities to deploy server software, load data, configure services, debug http traffic if there are problems. Data preparation (if standard interchange formats are being used) requires understanding of ETL using SQL queries and other techniques. Some understanding of XML and XML schema is required occasionally; GeoServer requires mapping from XML to database fields in an XML configuration. Someone with understanding of metadata content models and encoding is likely to be essential to get a catalog system working well.
* Outreach and marketing: personnel dedicated to the production and maintenance of documentation and educational materials, as well as face-to-face and online training programs.

Maintenance:

* Individuals dedicated to system management, including arrangement of meetings, maintenance of hardware, user help lines, and network operation.
* Software development and testing:

Components for service deployment may be necessary if off the shelf solutions do not meet all requirements; applications for service and conformance testing, development of metadata, and network monitoring.

Transition from current stovepipe data discovery and delivery systems to a loosely coupled, service based architecture can probably be done by large organizations like the USGS with the personnel and hardware they currently utilize, following whatever IT/hardware refresh cycles they currently use. Additions to hardware capacity in the form of servers, bandwidth, and online storage will be necessary to bring new data online, but this would be true no matter what approach to data delivery is adopted. The most significant additional investment will likely be in education of personnel who develop and deploy data and metadata services, in addition to human effort for data integration, documentation, and migration of existing data to new formats and delivery protocols. These investments will provide long-term return in staff capabilities if, as anticipated, these become widely accepted standard operating procedures. An additional return is the expected increase in utilization of the information resources and greater visibility for the agency providing data.

Maintain system providers/nodes

Even though each node will make their own choices regarding computer manufacture, operating system, network providers, etc., (another strength of a distributed network) and while there is currently no provision for ongoing federal support of individual nodes on the NGDS, it would be worth considering implementing a Service Level Agreement (SLA). SLAs are a common vehicle in the IT world for specifying expectations including system quality attributes such as interoperability, reliability, availability, recoverability, performance, integrity, confidentiality, etc. Suggested standards could potentially include some or all of the following (Allison, et. al., 2013):

* Minimum storage capacity and provisions for added growth of disk space
* Minimum network connectivity bandwidth
* Regular backup and maintenance procedures according to a schedule, with off-site storage of backups
* Codified disaster recovery plan with alternate facilities (For those organizations without an alternate backup facility, they could agree to serve as a backup facility with another node in exchange for the same)
* Physical system security precautions
* Software system security precautions
* User administration support (for data contributors to that specific node and/or to block inappropriate use from within their network domain)
* Commitment to participate in the NGDS using the standards and protocols established by the lead team

In an effort to fully meet the FOA requirements, external privately-licensed data may be required. For example, private entities own and maintain GIS layers of the U.S. Electric Grid. Licensing of this data should be the responsibility of the eventual home of NGDS. Costs for licenses will vary depending on the data. This may be one of the added features/functionality of paid subscription services (further discussed in the Managerial Sustainability section) for the NGDS to help maintain the system. In addition, certain data sets, such as the Power Grid, may require additional access control, those requiring authenticated user-IDs to log in to the system for maintaining security.

The inclusion of additional geospatial data is also a bonus to the distributed nature of the NGDS. Because of the project’s flexibility, it is scalable and transportable to other potential end users including the oil and gas exploration industry (particularly since a large percentage of the data in NGDS is related to oil and gas exploration), hydrology, and land-use and planning. By targeting additional end-users, NGDS is able to diversify its clientele base, including the critical data consumer clientele.

Finally, due to the systems scalability and transportability, additional partners and participants within NGDS and USGIN are possible. For example, NGDS was recently featured at the Canadian Geothermal Energy Association (CanGEA) Annual Meeting. Subsequently, the first Canadian Geothermal data set was submitted to the NGDS catalog. This isolates a potential issue with the name “National Geothermal Data System” and whether or not the inclusion of international data would warrant a name change for the system to the “Geothermal Data System” (additional discussion on this subject will occur in the next section).

An additional potential partner includes the State Department, which through a collaboration between USAID and the U.S. Energy Agency (USEA), is working to promote geothermal, particularly in the East African Rim, to provide business opportunities for U.S. companies. Preliminary conversations with State have commenced and will ensue during the sustainability phase of this project.

Furthermore, the Department of Defense has funded a collaboration community known as the Western Regional Partnership (WRP) to coordinate between 16 federal agencies in an effort to discuss land-use issues and encroachment on military facilities and wildlife protection. As part of the WRP project, more than 10,000 land-use data layers are being compiled. Metadata for the layers are being created and registered into the NGDS catalog and a series of datasets will be deployed as WMS, WFS. While progress on this partnership has been slowed (the initial working agreement was signed in late 2011), the current work plan includes deployment and connection with USGIN/NGDS. Since the DOD is expected to transfer this business model to the remaining regions in the US, there may be the opportunity to benefit from the partnerships in addition to some funding opportunities to help sustain USGIN, and by extension, the NGDS.

**Estimated Costs for Average Node Maintenance**

Currently, Arizona Geological Survey (AZGS) Nodes are budgeted at approximately $75,000 per year (originally $100,000 per year; however with the no-cost extension and slower initial spending, the rate is more in line with $75,000 per year) (Table 2). This includes personnel for approximately ½ time in order to assist with deployment of the services and initial server and licensing costs. For ongoing maintenance we estimate that this workload will reduce dramatically and move into maintenance mode with limited new data updates. Thus, we estimate a total of 1 hour per week for a informatics supervisor; 2 hours per week for an IT specialist with a 40 hour block for IT services, in case of emergencies, and approximately 4 weeks (or 160 hours) for a content manager. In addition, we estimate hardware costs to be approximately $2,000 per year for electricity, cooling, space, internet connection, and miscellaneous small part replacement (such as cabling, connectors, etc.); software licensing if using SQL server or ArcGIS server, although these rates are highly variable. Finally we have estimated through our existing Virtual Machines (VMs) and Amazon Web Services contracts for servers and storage associated with NGDS services that approximately $500 per year per Virtual Machine (VM) is needed plus an additional $600 per VM per year for cloud storage. The total base-cost per node is approximately $23,000 per year with existing servers.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Item | Description | Unit | Unit Cost | Total Cost |
| Supervisor | Approximately 1 hour per week | Hrs/yr | $65 | $3,120 |
| Technician | 8 hrs/month, intermittent | Hrs/yr | $50 | $6,800 |
| Content Manager | 160 hrs/year | Hrs/yr | $35 | $5,600 |
| Hardware incidentals | Electricity, cooling, space, etc. | /yr | $800 | $800 |
| Internet | Broad-band connectivity (as part of a larger network) | /yr | $100 | $100 |
| Software | Site license | /yr | $2,500 | $2,500 |
| Machines | Servers, storage – based on existing HP server running ~20 VMs, 5 year lifetime | Per VM/per yr | $500 (4 VMs) | $2,000 |
| Cloud Storage | Amazon Web Services, small instances (+/- 30%) | Per VM/per yr | $600 (4 VMs) | $2,400 |
| TOTAL |  |  |  | $23,320 |

Table . Yearly cost of AZGS data nodes.

One potential funding mechanism for the NGDS data hubs is to act as repository services to the federally funded science community or even federal agencies to help meet the new Open Data initiatives. This could be done through a series of “accredited” repositories where data is submitted, and reviewed for compliance with repository requirements.

**Education, Outreach, & Training (EOT)**

One of the key components for maintaining the system is ensuring that end-users have access to appropriate materials and training courses for the system. AZGS is currently engaged in an active Education, Outreach, & Training (EOT) component comprised of conference talks and demonstrations, and exhibit booth participation and key stakeholder events such as the Geothermal Resources Council (GRC) Annual Meeting and Geothermal Energy Association’s (GEA) Annual Expo, the Geological Society of America’s Annual Meeting, and the American Geophysical Union’s Fall Meeting. While this has been successful in engaging members of the broader community, additional targeting is required for specific end-users. AZGS is currently engaging the outreach team at the survey to review current user documentation and websites, as well as develop the final CKAN NIAB user documentation. In addition, a series of workshops for current data contributors and hubs are taking place to demystify interoperability and data services. Finally, an end user workshop for data users is being developed to be held in conjunction with the GRC/GEA in October of 2013. This workshop is designed to encourage data users from the industry to learn about the NGDS and being its use.

**Key findings for Operational Sustainability:**

* A distributed system’s weakness and strength is that it relies on multiple stakeholders for maintenance and management.
* Virtual servers offer rapid extensibility, less onsite maintenance, and overall lower costs than onsite hardware. In addition, they offer flexibility for rapid replication and disaster recovery. Thus, it is recommended that virtual machines be used for web service deployment and replication when at all possible.
* Long-term developer engagement should be encouraged by maintaining the developers forums, code-sharing repositories, and system documentation.
* Service Level Agreements (or Recommendations) should be prepared for additional (or potential) nodes.
* Licensing costs must be considered in the long-term maintenance of the system; offsets for these costs included authenticated user-IDs and subscription payments for added benefits.
* Extending the use to other user domains may encourage additional revenue
* In regard for international datasets, international nodes may consider implementing their own “National” Geothermal Data System or the system as a whole may consider changing its name to the Geothermal Data System with National nodes (although this may negate much of the existing publicity already established for the NGDS)

## Organization & Management

**Organization & Management of the System (Allison, et. al, 2013)**

As mentioned previously, there is no ‘one’ governance system that comprises the NGDS, nor is there one centralized responsible party. Rather, it is a loose collection of systems and organizations, each with their own methods of operation and management. USGIN is a framework of adopted standards, protocols, and practices, which weaves the NGDS components together into a federated distributed network. To help conceptualize this concept, the internet can be used as an analogy – there is no “owner” of the internet, rather an accepted set of standards and protocols necessary to participate. Perhaps the closest “owner” of the web is the W3C, or World Wide Web Corporation, an international community where member organizations and full-time staff, in conjunction with the public, work together to develop Web standards in order to lead the Web to its fullest potential. W3C is profiled in Appendix 1 as an organizational case study for potential implementation for NGDS or USGIN. Currently, NGDS is currently the largest ‘client’ of USGIN. To expand, and return to the World Wide Web example, NGDS (DOE) would be a member organization of USGIN (W3C), helping to develop standards and protocols for data sharing.

The next-phase of NGDS operation should be structured to eventually pursue revenue streams as necessary and appropriate including from grants and contracts (agencies, foundations, private sector), membership, fees for services (consulting, training, customization, ‘app’ development), repository services (data, services, apps, models, documents, multimedia), advertisements, fees for premier services or applications, subscriptions to value added services, licenses, contributions and donations, endowments, and sponsorships.

Who owns NGDS?

The question of who owns NGDS is not resolved, nor has it been discussed in a meaningful way. When the original Design-Build project at Boise State University was the only activity underway, it was not an issue. When the other projects were funded by DOE, the then Principal Investigator stated that “National Geothermal Data System” and “NGDS” could only be used by BSU. Subsequently, the BSU project referred to project materials being copyrighted by NGDS. DOE project personnel raised questions about that. Materials produced by the Arizona Geological Survey under contract to BSU were included in the copyright claims, raising objections by AZGS. The copyright claims were withdrawn, but this helped bring forward an issue that had been largely ignored up to that point.

Before we can decide who owns NGDS, we need to understand what NGDS is. It is set of Web services provided free of charge to end-users, built upon a collection of open source standards, specifications, procedures, and protocols, assembled using the US Geoscience Information Network.

The data being provided via NGDS is nominally owned by each of the data providers (not withstanding data in the Geothermal Data Repository) who were funded to digitize and serve them, albeit according to system requirements.

USGIN is being enhanced and made operational using NGDS funding. The regional server hubs managed by the AZGS/AASG project did purchase some hardware and software to store, serve, and backup data in NGDS.

The ‘backend’ of NGDS is decoupled from the ‘front end’ or user interface. So, while there may be an official NGDS portal, it is not necessarily the only portal to the system. In fact, we have promoted the idea of anyone building a value-added portal to NGDS data and services, similar to anyone building tools to access the World Wide Web (e.g. Yahoo, Internet Explorer, Google search, etc).

Thus, NGSD is comprised of a set of mostly intangibles – a virtual collection of standards and agreements to use them, among data providers. The value of NGDS increases as more data and software (“app”) providers adopt NGDS standards. In our current nascent stage of development of interoperability, there are many options that may eventually become a national system of standards, leaving NGDS as a niche functionality or curiosity. The agreements with the NGDS data/service providers are not exclusive. Anyone can adopt the same or similar standards and procedures, tap into the free online data stream and effectively deliver NGDS themselves. However, the success of the AASG part of the projects is that the State Geological Surveys have a century plus legacy of trust and cooperation, that allowed that group to function as an established team in a way that 50 independent unaffiliated organizations could not.

The owner of NGDS will be responsible for maintaining the flow of data and services from providers, keeping the system operational and updated, managing the system, and ensuring adequate funding.

Our position is that DOE ‘owns’ NGDS as a brand and a service, and those are valuable commodities. DOE can manage that themselves, contract it out, put it up for bid to manage, license or sell it.

Transition period after ARRA funding

It is anticipated that NGDS will not be in full independent sustainable mode immediately at the end of the DOE ARRA funding in May, 2014. This is due to a number of factors including ambiguity about the ownership and functionality of NGDS prior to the redirection of the Design-Build project. Thus, the more serious sustainability planning began perhaps a year later than it might have.

In addition, we do not have a fully functioning system that can be demonstrated or fully “marketed” to potential clients, sponsors, or other potential funding sources. The system is not intended to be fully operational until then; the user interface which will more effectively demonstrate the capabilities of the system will not be available until Fall, 2013; and integration of the other NGDS projects (e.g., those at SMU and USGS, plus the Geothermal Data Repository) into the system is underway with somewhat uncertain timelines.

As a result, we expect there will be a transition period from DOE sponsorship to full sustainability. The period and scope of that transition is dependent on the decisions made in the sustainability planning process and success in achieving them.

Internationalization of NGDS – global GDS, national GDS’s

The growing interest from the international sector argues for at least considering changing the name from National Geothermal Data System to just Geothermal Data System to make it more universal in scope and attractive to non-US entities.

Alternatively, we could move towards a federated system of many national GDS’, with NGDS being the US component, and country-named companions (e.g., Canadian Geothermal Data System, etc).

A global network also offers other countries the ability to manage their own networks rather than deferring to US control, as long as there is agreement among the participants in adoption and maintenance of standards and procedures. The globalization of cyberinfrastructure governance and system architecture is well underway and we can draw on those activities for exemplars.

Organizational status

NDGS could be managed and run by DOE or a national laboratory as an entirely federal activity. It can be spun off as independent stand-alone entity, responsible for its own survival. Alternatively, it can be housed in an existing body during a transitional period analogous to an incubator start-up, or on an ongoing basis (permanently or renewed/rebidded periodically).

We have been reviewing a number of bodies with differing business and organizational models (see Appendix I).

*Stand-alone organizations* offer the greatest independence and likely adherence to the organizations vision, goals, and missions. It also demands more challenges to start up and establish. An organization dedicated to the longevity and success of NGDS won’t be abandoning it for any number of reasons as others (below) might.

Issues we will have to deal with in setting up a stand-alone organization include

* Where to legally locate
* Where to physically house
  + Virtual vs centralized
* Funding needed to form a non-profit
* Management structure
* Administrative support (accounting, payroll, facilities, personnel)
* Advisory Committees or Board (management, technical, scientific)
  + Board of Directors
  + Executive functions
  + Staffing

*Incubators (transitional host)* generally have a vested interest in seeing the organization develop and thrive but cannot accept the costs or risks of long-term or permanent commitment to maintain the functions.

*Parent organizations* could come from any of societal sectorsegments, in which NGDS could find a home as a program within an established organization that ishas compatible, including non-profits, academia, government, or industry. Non-profit cyberinfrastructure organizations organization (e.g., ESIP, OGC, USGIN) offer the value of supporting technical infrastructure. An advantage is the inherent emphasis on using the data integration framework in other domains, broadening the base of users and supporters. The downside is being run by groups that are more focused on technology than user needs and priorities.

Non-profit geothermal organizations (e.g., GRC, GEA) bring the advantage of focusing on end-user needs which are often secondary to computer developers. A concern might be whether the NGDS would compete or be seen as competing for resources within the organization, although many professional societies have or are developing digital resources that they market (e.g., AAPG Datapages, SEG OneMine). Those however are fee-based approaches that may not accommodate the free distribution of data and services in NGDS now.

Universities can be excellent sites for start-ups because of the access to technology, infrastructure, and highly educated workforce. On the other hand, tech start-ups in universities typically spin off to have more financial independence as soon as they can. The NGDS mission as an operational system rather than a research project also argues against a university home in general although there may

Non-profit geothermal organization (e.g., GRC, GEA)Agencies (state or federal) seem unlikely long term hosts for NGDS. The vagaries of budgets and political priorities provide a level of uncertainty and risk higher than desired. It is usually difficult for state agencies to be engaged in leadership roles in projects that are national in scope. Concerns that state funds are being used to subsidize extra-state activities put such projects at risk of being shut down. Hosting NGDS in a federal agency presumes that agency will fund it. It would be difficult for a federal program to pursue many of the revenue streams we identify.

Commercial hosting of NGDS would only seem to work if the company can make a profit off of it. That could still allow free access to the NGDS data and services – after all, Google gives away maps, images, and many other services to users for free, with an ad-based and fee for premium services model. In fact, we consider some of those same revenue streams as possibilities for NGDS.

A risk with a commercial host for NGDS is the potential for the company to shut it down at any time with little or no notice if it loses money or no longer fits with the company’s business strategy. That has happened with Google products as well as many others.

Business structure

[The following is adapted from the USGIN sustainability study completed in 2011.]

Depending on decisions about the activities and responsibilities of an NGSD (or USGIN) organization established in a governance charter, a defined level of financial support will be necessary and should be detailed in a business plan. Based on a cursory survey of existing organizations with intent similar to USGIN and NGDS, conducted by the USGIN Working Group, they recommended that a non-profit foundation model may be the most appropriate. This would define NGDS or parent organization as a legal entity that could enter into contracts, receive funding, pay salary, and make grants of funding. A formal legal entity could support a small number of full time employees who look after day-to-day operations, planning, and logistics for community events. An important issue to be determined by the governance authority is whether there should be a stand-alone organization, or if it is more effective to form a group within the framework of some other existing organization. The implications for a separate business entity or affiliation with an existing organization need to be explored in the context of financial and intellectual property considerations that may result from the government agency nature of most geological surveys.

The business structure to support a community geoinformatics development process should promote grass root nucleation of interest groups that are the core of development efforts. Selection of groups to be nurtured by available funding should be based on community assessment of priorities and user demand. Resource allocation by government agencies in support of these activities must be tempered by programmatic priorities. Loosely knit organizations (Appendix I) provide organizational models, and the NGDS community will progress most efficiently by utilizing these existing groups to gather input and effort from the communities they attract.

Because many geological surveys have data archive and dissemination functions as part of their portfolio, some support for NGDS might be built into their operating expenses and overhead as users come to expect access to data resources through standardized network interfaces. Sharing of resources and reuse of components through NGDS (and USGIN) will reduce the cost of these activities. Wide adoption of similar software, protocols and practices increases the number of stakeholders with an interest in supporting NGDS and USGIN. Services and data streams deployed by the geological surveys must have sufficient value to the user community to support either a pay-for-use model, or to motivate continued public funding if the system it to be viable in the long term. NGDS can be seen as a metaphor—it is building an online version of the bricks and mortar libraries that historically have been the anchor of knowledge preservation and access for the geosciences.

The American Recovery and Reinvestment Act of 2009 was the primary means through which the various elements of the NGDS were created, with oversight provided by the EERE’s Geothermal Technologies Office. The awards expire as early as September 2013, in the case of the SMU project, with no confirmation of continued federal funding. Various business models and plans used by other nationwide geo-information data and service providers dedicated to other specialties (seismology, meteorology, hydrology, etc.) are under review to determine which methods have been successful in long term funding support for the inherent costs of maintaining a system of the size of the NGDS and its major nodes. Then National Science Foundation supports a number of these other efforts, but there are other models worth consideration involving commercialization of some portions of the system.

While it is unlikely that something built primarily by government funding would be entirely privatized, establishment of a non-profit entity would open up possibilities for membership support, advertising income, custom services available for a fee, as well as support from grants and/or contracts.

Data being added to NGDS are not strictly geothermal only but data relevant to geothermal exploration and development, including vast amounts of oil and gas and groundwater wells and other data. These are relevant to a broader base of users. By diversifying the client base of NGDS to other users and other fields, the cost of maintaining core infrastructure can be spread across an array of stakeholders, many of which are larger than the DOE Geothermal program. It is presumed that NGDS will continue to provide free and open access to its data resources.

The next-phase NGDS operation should be structured to eventually pursue revenue streams as necessary from grants and contracts (agencies, foundations, private sector), membership, fees for services (consulting, training, customization, apps), repository services (data, services, apps, models, documents, multimedia), advertisements, fees for premier services or applications, subscriptions to value added services, licenses, contributions and donations, endowments, and sponsorships.

Additionally, several of the institutions contributing data to the NGDS are involved in research that crosses U.S. boundaries and interest has been expressed in how to leverage the U.S. NGDS efforts for global impact. This is another area that could open up possibility for a source of financial support of the NGDS.

As of this writing, the final determination as to who “owns” the NGDS or how it should be managed after the ARRA funding has expired has not been finalized. Our preliminary concept is to treat NGDS as a DOE-owned asset that can be managed by an external entity under contract or from a grant from DOE (although as stated in the assumptions, this is not expected). NGDS is run on the US Geoscience Information Network data integration framework of standards, protocols, and practices. USGIN is expected to be established as a non-profit entity and could maintain NGDS as one of a number of loosely coupled client applications on behalf of DOE. Alternately, NGDS could operate independently or within another organization such as a professional society or industry trade association, and use USGIN just like other open source software (e.g. CKAN), or code repositories (e.g. GitHub) and standards (e.g. OGC, ISO).

Clients and Markets

One of the first steps as the NGDS becomes more operational is to assess who is using the data and services and what value it brings to them. The community feedback thus far has focused on designing and developing the user interface rather than the system as a whole. We also need to assess the value to DOE as the major sponsor and to the U.S. from economic and national security perspectives.

We have presumed the value to the geothermal energy industry, to government agencies, and to science researchers, in the venues we have chosen to demonstrate and exhibit the system.

Since the beginning of the project we have been reaching out to other users and providers (WRP, Energistics, National Science Foundation, State Department, Microsoft Research, National Data Repositories, and so on) to broaden the base of users and providers to make USGIN/NGDS the leader in digital interoperability and the default system for new adopters. The more widely used the system is, the more critical it is to users that it be sustained and that they have a say in it.

We believe the NGDS framework can be marketed to the natural resources industries (energy and minerals), government agencies (especially in light of the White House Executive Order on public access to data), academia (as evidenced by the direction NSF EarthCube is moving), industry (who want access to the free data stream from NGDS to make value added products), and internationally (via many activities and connections underway already and emerging ones).

Costs of operations

More detailed cost estimates are needed to determine minimum and optimal costs for these primary tasks:

NGDS Maintenance (status quo)

NGDS Enhanced (more sophisticated user interface and visualization, added nodes, data and

services)

USGIN Maintenance

Business Development

Revenue sources

We have done extensive research on existing organizational and business models for entities that are analogous in one or more ways to NGDS (see case studies in Appendix I). Among the topics we investigated are the revenue sources, particularly for non-profit groups. We have identified an extensive (but probably not exhaustive) list of ways to fund NGDS, including:

* Grants and contracts (agencies, foundations, private sector)
* Membership (of the organization, of the governing council)
* Fees for services (consulting, training, customization, digitizing, apps, node setup)
* Repository services (for data, services, apps, models, documents, multimedia)
* Services and resources (for permitting agencies)
* Ads (it works for Google)
* Fees (for premier services or applications)
* Subscriptions
* Licenses (to use the NGDS brand or logo; to access some element of user data)
* Contributions-Donations
* Endowment
* Sponsors/underwriters (get acknowledgement on site and in materials)
* For-profit companion/subsidiary (spin-offs to commercialize some service or product)
* Conferences, workshops

Some of these are contradictory or contrary to the free access currently in place. Not all (or even most) of them may ever be implemented, nor would they all run concurrently. Each will take time and resources to develop, so they need to be prioritized in terms of risk and return on investment.

We have de facto been focusing on grants and contracts since there are existing funding mechanisms that we simply respond to. But we are also actively building partnerships with users who are potential sponsors, and offering support and preparing training materials for anyone who wants to tap into the NGDS data stream, or learn how to contribute data, develop apps, or become a node.

# conclusions

The future of NGDS depends on taking measured steps to build this initiative and see it succeed. A summary of recommended actions from this document include:

* Develop and deploy a marketing strategy to increase engagement of the State Surveys, other data and service providers, and stakeholders to cement a strong user base supporting NGDS (and USGIN).
* Establish a governance structure led by a working group with representatives from NGDS participants and stakeholders.
* Decide if NGDS should be established as an independent non-profit 501(c)(3) foundation or a subdivision of an existing organization
* Conduct a comparative analysis of preferred sustainability models and select one to implement that will provide a conduit for financial resources to support infrastructure, a small (2-3 person?) staff, and servers for system infrastructure
* Inventory the State and Federal Geological Survey information architecture landscape to identify and prioritize use cases, and to identify capabilities and expertise to develop prototypes and testbed opportunities.
* Work to promote consistent best practices.
* Engage with domain science and computer science communities to promote interests of NGDS/USGIN in national cyber-infrastructure development.

# recommended Solutions

We propose preparing a formal business model based on NGDS being licensed by DOE to a newly formed non-profit USGIN for some period to get them both established financially and functionally. The business model should be vetted with the broad geothermal community and other potential sponsors, supporters, and clients of USGIN and NGDS services and data.

Client of USGIN

We propose that NGDS and USGIN be jointly developed at least during the initial phase of making

NGDS sustainable. NGDS is developed on the USGIN framework which is still emerging and maturing. NGDS serves as the largest client of USGIN services and is driving much of the USGIN technical development and data provider network (i.e., through AASG). Until both USGIN and NGDS reach full operational status, it is prudent to have them closely tied.

Within a few years when the systems are each more of an operational commodity we anticipate that NGDS could be easily decoupled from USGIN and run independently, if there was reason or desire to do that.

USGIN Non-profit

The rationale and justifications for taking USGIN to non-profit status (Hutchison and Richard, 2011) are equally applicable to NGDS. This structure opens multiple opportunities for adoption and support by government agencies, funding from agencies and foundations, and even support from industry.

Non-profit status does not prevent the setting up of a for-profit spin-off at a later time for commercialization of products or services if warranted.

Broad-based stakeholder management

Engagement of other end users and data contributors of the system will be facilitated by providing those entities opportunities for meaningful participation in the governance of the system. This may be through membership, advisory committees, boards, or comparable mechanisms.

Re-brand as GDS

NGDS is funded by US DOE GTO to catalyze and facilitate geothermal energy exploration and development in the U.S. However, there is a growing international interest in adopting not only the NGDS technical approach but NGDS itself. The Canadian Geothermal Energy Association has test demonstration data set served through NGDS and is reviewing what is required to establish a full service node on NGDS serving Canadian data.

The U.S. State Department proposed that NGDS be deployed to Latin America and East Africa, the latter in cooperation with the recently established East African Geothermal Program financed by USAID through the US Energy Association. Lawrence Berkeley Lab is developing a 10 week long geothermal training course for the EAGP to begin in Summer, 2013 and asked NGDS to provide a week long segment on data management.

As NGDS provides more data and services and becomes more widely known and used, we anticipate that other adopters will similarly come forward. Renaming it to the “Geothermal Data System” would demonstrate the global capabilities of the system, reduce perceptions that it is only for US providers and users, thus broadening the user base and potential supporters.

Individual nodes (e.g., countries) could brand their deployments of the GDS as [country name]GDS such as “Canadian Geothermal Data System (CGDS)” within the larger global network, effectively as a franchise.

Diversified client base

The NGDS targeted clientele is the U.S. geothermal industry and most of the NGDS data and services focus on the exploration and development parts of the industry. While that is a substantial economic entity it is still tiny compared to the potential natural resources, geoscience, and geospatial customer bases for NGDS data and services.

Diversified funding streams

We believe that a multifaceted approach to funding NGDS will be required. Not all the funding mechanisms we list will be viable, and in fact, some are contradictory or exclusionary. Nor will all of the viable mechanisms need to be or can be implemented initially. Our situation is not dissimilar to that of many Web-based businesses. They use a variety of business models, they experiment with them, and add or drop revenue streams as part of a highly entrepreneurial and evolutionary process. We need to emulate that same flexibility and innovation.

# Implementation Plan

Decision Process

The decisions on how to proceed will be made by US DOE GTO, primarily in consultation with BSU and AZGS, the latter acting on behalf of AASG. If affiliation with USGIN is desired, AASG and USGS need to be involved.

Responsibilities

Creation of a non-profit USGIN will be led by AZGS for AASG. It is the intent of AASG and USGS for this to happen, regardless of what DOE decides on NGDS.

If it is agreed that NGDS is owned by DOE, then they will need to determine how to delegate management of it via a contract, license, etc.

Alternatively, if DOE determines that NGDS is a product of BSU or BSU and the other NGDS project organizations, those entities will be solely responsible for decision making. However, given DOE’s substantial financial and intellectual investment in NGDS and anticipated compelling interest in seeing it succeed, it is expected that DOE will continue to have significant input (and investment) into operations.

Timeline

Decisions on the four areas of sustainable need to be finalized as quickly as possible in order to start the processes for technical, organizational, operational, and financial sustainability. The concern is that we cannot achieve these by the time the DOE funded projects end in spring 2014. There will be some transition period after April 2014 but we cannot predict the length and scope of that until the sustainability plan is adopted.

Resources

The AASG project has set aside about $250,000 in funding that was returned from project subcontractors who mostly could not complete the original scopes of work. We propose allocating that to implementing the business model for NGDS. This includes creation of informational, promotional, and educational materials, and hiring of a development manager, possibly based in Washington DC area, to line up additional end-users and system participants. Care will be taken to ensure that no federal funds are used in seeking additional federal financial support.

# Acknowledgements and Disclaimer

This material is based upon work supported by the Department of Energy, Office of Energy Efficiency and Renewable Energy, Geothermal Technologies Office, under Award Numbers [*for BSU and AZGS]*

This report was prepared as an account of work sponsored by an agency of the United States Government.  Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights.  Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof.  The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

# References

Arizona Geological Survey, “Setting up Hyper-V 2012 Replication on Workgroup Servers: a guide,” 5/10/2013, available at: <https://github.com/usgin/Workshops/blob/master/WesternRegionalWorkshop/Hyper-V%202012%20Replication%20on%20Workgroup%20Servers.docx>, soon to be available on [www.geothermaldata.org](http://www.geothermaldata.org) under “Tutorials.”

Allison, M. Lee, Tim Ahern, David Arctur, Jim Bowring, Gary Crane, Cecelia DeLuca, Geoffrey Fox, Carroll Hood, Hannes Leetaru, Kerstin Lehnert, Chris MacDermaid, Mohan Ramamurthy, Erin Robinson, Ilya Zaslavsky, Genevieve Pearthree, Kim Patten, and Debra Winstead, 2012, “EarthCube Governance Roadmap: Documentation, Research, and Recommendations, Version 2.0.,” 237p, <http://www.nsf.gov/geo/earthcube/docs/EarthCubeGovernanceRoadmap.pdf>

Allison, M. Lee, Tim Ahern, David Arctur, Jim Bowring, Gary Crane, Cecelia DeLuca, Geoffrey Fox, George Percivall, Carroll Hood, Hannes Leetaru, Kerstin Lehnert, Chris MacDermaid, Mohan Ramamurthy, Erin Robinson, Ilya Zaslavsky, Genevieve Pearthree, Kim Patten, Jay Pearlman, and Joe Cutcher-Gershenfeld, 2012, “EarthCube Governance Framework: A Proposal to the Community; Documentation, Research, and Recommendations,” 200p. <http://www.nsf.gov/geo/earthcube/docs/EarthCubeGovernanceFramework.pdf>

Allison, M. Lee, Anderson, Arlene, Chickering, Cathy, Patten, Kim, Richard, Stephen M., Tanikella, Rajanikanth, “Sustaining the National Geothermal Data System: Considerations for a System Wide Approach and Node Maintenance,” Geothermal Research Council, 2013 Annual Meeting and Conference, DRAFT paper, submitted for consideration.

Blackman, Harold, Boyd, Suzanne, Patten, Kim, Zheng, Sam, “NGDS User Center Design: Meeting the Needs of the Community,” 38th Stanford Geothermal Conference Proceedings, Stanford, CA, February 11-13, 2013.

Blackwell, David D., Chickering, Cathy, Hughes, Allen, Kline, Don, Moerchen, Fabian, Tanikella, Rajanikanth, “Sustainability Plan for the SMU Node of the National Geothermal Data System,” Version 1.2, August 30, 2012.

Clark, Ryan J., Kuhmuench, Christoph, Richard, Stephen M., “Developing the National Geothermal Data System Adoption for Domestic & International Data Development,” 38th Stanford Geothermal Workshop Proceedings, Stanford, CA, February 11-13, 2013.

Deloitte, for the U.S. Department of Energy, Geothermal Technologies Program, “Geothermal Risk Mitigation Strategies Report,” <http://www1.eere.energy.gov/geothermal/pdfs/geothermal_risk_mitigation.pdf>, February 15, 2008

Geothermal Data Consortium, “National Geothermal Database,” proposal narrative, in response to Funding Opportunity Announcement DE-P36-08GO98020, January 2009.

Edwards, Paul, Steven Jackson, Geoffrey Bowker, and Cory Knobel, “Understanding Infrastructure: Dynamics, Tensions, and Design - Report of a Workshop on “History & Theory of Infrastructure: Lessons for New Scientific Cyberinfrastructures,” 2007. <http://cohesion.rice.edu/Conferences/Hewlett/emplibrary/UI_Final_Report.pdf>

European Geological Data Infrastructure (EGDI) - [Hazel Napier]

Executive Office of the President, Office of Management and Budget, Memorandum for the Heads of Executive Departments and Agencies, M-13-13, “Open Data Policy – Managing Information as an Asset,” May 9, 2013.

Hutchison, Vivan, and Stephen Richard, 2011, “Recommendations for the Future of the U.S. Geoscience Information Network,” Joint report of the Association of American State Geologists and the U.S. Geological Survey, 11p.

Hutchison, Vivan, and Stephen Richard, 2011, “Recommendations for the Future of the U.S. Geoscience Information Network,” Joint report of the Association of American State Geologists and the U.S. Geological Survey, 11p (Attached)

Is a geospatial user fee the answer to creating and maintaining the U.S. National Spatial Data Infrastructure? <http://www.directionsmag.com/podcasts/is-a-geospatial-user-fee-the-answer-to-creating-and-maintaining-the-u./318049>

National Geothermal Data System Architecture, Design, & Testing (NGDS System Architecture), “Software Requirements Specification, NGDS” *Working Document*, Version 2.7, November 5, 2012.

National Research Council, The Future of Scientific Knowledge Discovery in Open Networked Environments: Summary of a Workshop, National Research Council, Washington, DC, <http://www.nap.edu/catalog.php?record_id=18258>

National Research Council, The Case for International Sharing of Scientific Data: A Focus on Developing Countries: Proceedings of a Symposium, NRC

National Research Council, The Future of Scientific Knowledge Discovery in Open Networked Environments: Summary of a Workshop, NRC

Standards and Infrastructure for Innovation Data Exchange <http://www.sciencemag.org/content/338/6104/196.full>, Science 12 October 2012: Vol. 338 no. 6104 pp. 196-197, DOI: 10.1126/science.1221840

U.S. Department of Energy, Geothermal Technologies Program, “Funding Opportunity Announcement: DE-PS36-08GO98020, National Geothermal Database” November 20, 2008.

Allison, M. L., & Dickinson, T. L. (2008). *A workshop on the role of state geological surveys and U.S. Geological Survey in a geological information system for the nation.* Tucson, AZ: Arizona Geological Survey.

Ferguson, W. (2013, February 25). *New Geothermal Data System Could Open Up Clean-Energy Reserves.* Retrieved May 1, 2013, from Scientific American - Digial: Permanent Address: http://www.scientificamerican.com/article.cfm?id=new-geothermal-data-system

Holdren, J. P. (2013). *Increasing Access to the Results of Federally Funded Scientific Research.* Washington, D.C.: Executive Office of the President, Office of Science and Technology Policy.

IBM Corporation. (2012, May). *Crossing the sustainablity chasm: Strategies and tactics to achieve sustainability goals.* White Paper, IBM Corporation, IBM Software, Somners, NY.

International Energy Agency. (2011). *Technology Roadmap: Geothermal Heat and Power.* Paris, France: International Energy Agency.

Richard, S. M. (2013). *USGIN White Paper - in development.* Tucson: Arizona Geological Survey.

U.S. Department of Energy, Geothermal Technologies Office. (n.d.). *About the Geothermal Technologies Office*. Retrieved May 3, 2013, from Geothermal Technologies Office: http://www1.eere.energy.gov/geothermal/about.html

U.S. DOE, Office of Energy Efficiency and Renewable Energy. (2011). *Geothermal Technologies Program Blue Ribbon Panel Recommendations.* Washington, D.C.: U.S. DOE.

The White House, Office of the Press Secretary, “Executive Order – Making Open and Machine Readable the New Default for Government Information,” May 9, 2013, <http://www.whitehouse.gov/the-press-office/2013/05/09/executive-order-making-open-and-machine-readable-new-default-government->

## Attachments and Links

Hutchison, Vivan, and Stephen Richard, 2011, “Recommendations for the Future of the U.S. Geoscience Information Network,” Joint report of the Association of American State Geologists and the U.S. Geological Survey, 11p

USGIN White Paper - [USGINWhitePaper.docx](https://github.com/usgin/usginspecs/blob/master/USGINWhitePaper.docx)

USGIN Metadata Requirements - [MetadataRecommendationsGeoscienceResource](https://github.com/usgin/usginspecs/blob/master/MetadataRecommendationsGeoscienceResources_v1.2DRAFT.docx)s

USGIN Identifier Scheme and Web Architecture - [USGIN\_URI\_Scheme\_1.1.doc](https://github.com/usgin/usginspecs/blob/master/USGIN_URI_Scheme_1.1.doc)

USGIN Information Exchange - [USGIN information exchanges.docx](https://github.com/usgin/usginspecs/blob/master/USGIN%20information%20exchanges.docx)

USGIN Guidance for Setting Up a New Information Exchange - [DefineNewInformationExchange.docx](https://github.com/usgin/usginspecs/blob/master/DefineNewInformationExchange.docx)

USGIN Standardization for Data Links [MetadataAsHypermediaApp.docx](https://github.com/usgin/usginspecs/blob/master/MetadataAsHypermediaApp.docx)

USGIN Standardization for Linking Information Resources Using Web Architecture - [MachineActionableLinksSummary.docx](https://github.com/usgin/usginspecs/blob/master/MachineActionableLinksSummary.docx)

USGIN Services Information and Announcements Feed - [USGIN services information and announcements feed.docx](https://github.com/usgin/usginspecs/blob/master/USGIN%20services%20information%20and%20announcements%20feed.docx)  (outlines content for an RSS feed that users would subscribe to for announcements and information about network service operations (down time, service sunsetting, new services…))

USGIN Profile for ISO Metadata - [USGIN\_ISO\_Metadata.doc](https://github.com/usgin/usginspecs/blob/master/USGIN_ISO_Metadata.doc)

# Appendix I: Case Studies

The following case studies were selected in order to provide analysis on governance and funding structures of existing data-related organizations. For each organization we provide 1) a general background of the organization’s mission and membership (if applicable), 2) the funding mechanisms in place, 3) the current governance framework and/or organizational model, and 4) staffing of the organization (either paid, volunteer, or a combination).

Case studies include:

* Domain science data centers: CUAHSI (Hydrology), IRIS (Seismology), Unidata (Atmospheric), IGSN (Physical Samples), OpenTopography (Topographical Data), OneGeology
* Information Technology: W3C Consortium (World Wide Web), DataONE (Environmental Data and Library Sciences), Apache Software Foundation (Open Source Software), DPN (Repositories), OpenDAP, IETF, DataBasin, DataCite, Open Knowledge Framework
* Federations: Earth Science Information Partners (ESIP – primarily serving the NASA/NOAA community) and Open Geospatial Consortium (OGC – geospatial standards)

### Domain Science Groups: CUAHSI, IRIS, Unidata, IGSN, OpenTopo

#### Consortium of Universities for the Advancement of Hydrologic Science (CUAHSI)

**General Background.** The Consortium of Universities for the Advancement of Hydrologic Science, Inc. (CUAHSI) is a 501 (c)3 research organization that represents over 130 U.S. universities and international hydrologic-science organizations. Its mission is to enable “the university water science community to advance the understanding of the central role of water to life, Earth, and society.”[[1]](#footnote-2) It is composed of hydrologic databases. CUAHSI supports its community to “advance water science and to improve societal well-being by:

* Developing, supporting, and operating research infrastructure;
* Improving and promoting access to data, information and models;
* Articulating and advocating priorities for community level water-related research and observations
* Facilitating interactions among the diverse water research community
* Promoting interdisciplinary education centered in water science
* Translating scientific advancements into effective tools for water management and policy.”[[2]](#footnote-3)

CUAHSI was incorporated in 2001 by 33 founding members, following more than 15 years of discussions, and meetings. CUAHSI was initially funded via an NSF proposal that lasted two years to plan a phased implementation.

**Funding.** CUAHSI is funded by the National Science Foundation.

**Governance. “**CUAHSI is governed by a [Board of Directors](http://www.cuahsi.org/BoardofDirectors.aspx), elected by and from the [Membership](http://www.cuahsi.org/Membership.aspx) and managed by the Executive Committee of the Board and its Officers. Members of CUAHSI's six [Standing Committees](http://www.cuahsi.org/StandingCommittees.aspx) are leading scientists and educators within the water science community, and provide community oversight and input into CUAHSI activities. CUAHSI Senior and Science staff are responsible for executing projects, providing services, and other day to day activities. Currently, over 50 scientists from both member and non-member institutions are actively serving the water science community on the CUAHSI Board and Standing Committees.”[[3]](#footnote-4)

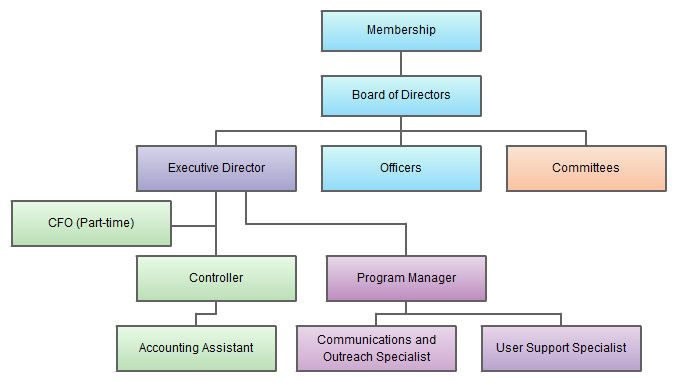


Figure . Cuahsi Leadership Structure. Source: “Governance,” Consortium of Universities for the Advancement of Hydrologic Science, Inc.

The CUAHSI governance framework is guided by a set of by-laws (<http://www.cuahsi.org/by-laws.html>).

**Staff.** CUAHSI Committees and a team of seven paid staff (distributed between the Program Headquarters [Boston] and the Administrative Headquarters [Washington DC]), carry out CUAHSI activities (Figure 1).

**Sustainability.** CUAHSI has 5-Year Strategic Plan, *Water in a Dynamic Planet: A Five Year Strategic Plan for Water Science*, which initiates an ongoing process of community review and input for all CUAHSI - led and CUAHSI - affiliated activities (<http://www.cuahsi.org/docs/stratplan/CUAHSI-5yr-StrategicPlan.pdf>). CUAHSI also has an Implementation Plan: 2013-2017 (<http://www.cuahsi.org/pageFiles/ImplementationPlan2013-2017.pdf>), that contains tactical decisions that focus CUAHSI's efforts and provide an approach towards achieving the vision of the *Strategic Plan.*

#### Incorporated Research Institutions for Seismology (IRIS)

**General Background.** The Incorporated Research Institutions for Seismology (IRIS) is a 501(c)(3) non-profit consortium of research institutions [more than 100 universities and others] founded in 1984 to develop scientific facilities, distribute data, and promote research. “IRIS programs contribute to scholarly research, education, earthquake hazard mitigation, and verification of the Comprehensive Nuclear-Test-Ban Treaty.”[[4]](#footnote-5) In addition, IRIS functions as an interface between the scientific community, funding agencies and IRIS programs, and facilities are multi-user and multi-use. The IRIS mission is to do the following:

* “Facilitate and conduct geophysical investigations of seismic sources and Earth properties using seismic and other geophysical methods.
* Promote exchange of geophysical data and knowledge, through use of standards for network operations, data formats, and exchange protocols, and through pursuing policies of free and unrestricted data access.
* Foster cooperation among IRIS members, affiliates, and other organizations in order to advance geophysical research and convey benefits from geophysical progress to all of humanity.” [[5]](#footnote-6)

**Funding.** IRIS is a 501(c)(3) nonprofit organization incorporated in the state of Delaware with its primary headquarters located in Washington, DC. IRIS was originally funded with support from the National Science Foundation. Every 5 years, IRIS must submit a new funding proposal to NSF, and IRIS members are encouraged to contribute to this proposal. The IRIS 2011-2013 proposal “Facilitating New Discoveries In Seismology and Exploring The Earth: The Next Decade” was successfully reviewed and forms the basis for a new Cooperative Agreement between IRIS and NSF to continue operation of the IRIS core programs. The new agreement runs through September 2013, when a new agreement is anticipated to support the merged operation of the core programs and EarthScope/USArray.

**Governance.** The IRIS governance framework establishes community ownership of instrumentation resources and data, and the framework is designed to allow IRIS to adapt to changes in technology and in the growth of the Consortium. IRIS’s community governance model provides a framework for decision-making that preserves community oversight of IRIS facilities. IRIS is governed by Consortium members (114 US research universities and research institutions – see Figure 2 below). Specifically, “The IRIS governance and management structure is an interface among the scientific community, funding agencies, and IRIS programs. The structure is designed to ensure close involvement of the research community in the development of IRIS facilities, focus scientific talent on common objectives, encourage broad participation, and effectively manage IRIS programs. Each year, over 50 scientists from member institutions participate in IRIS management through its [Board of Directors](http://www.iris.edu/hq/about_iris/governance/bod), eight regular committees, and ad hoc advisory groups. Standing committees maintain community oversight while program managers are responsible for day-to-day activities. Committee scientists work with a professional staff at IRIS…Representatives of the full set of member institutions elect the nine-member [Board of Directors](http://www.iris.edu/hq/about_iris/governance/bod).”[[6]](#footnote-7)

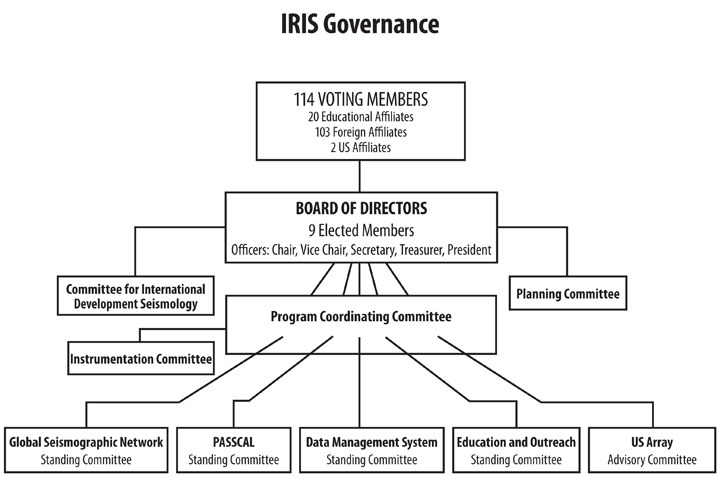


Figure . IRIS Governance, Source: “About IRIS,” last modified 2012, <http://www.iris.edu/hq/about_iris>.

Standing Committees and Board of Directors and Subcommittees make decisions regarding planning and operations. In addition, “Educational and not-for-profit institutions chartered in the United States, with a major commitment to research in seismology and related fields, may become [Voting Members](http://www.iris.edu/hq/about_iris/membership/member_institutions). Research institutions and other not-for-profit organizations both inside and outside the United States engaged in seismological research and development, which may not otherwise qualify for IRIS membership, may be elected [Educational Affiliates](http://www.iris.edu/hq/about_iris/membership/educational_affiliates), [US affiliates](http://www.iris.edu/hq/about_iris/membership/us_affiliates), or [Foreign Affiliates](http://www.iris.edu/hq/about_iris/membership/foreign_affiliates).”[[7]](#footnote-8) IRIS also employs workshops, annual meetings, symposia, and newsletters to promote on-going interactions with scientists and member institutions and to receive feedback. These mechanisms allow the research community to communicate its evolving needs to funding agencies.

**Staff:** IRIS is led by the President, Director of Planning, Director of Project Administration, Director of Finance and Administration, and the five program managers, in addition to more than 50 staff members distributed throughout IRIS headquarters, Data Management Center, PASSCAL, Education and Outreach, Global Seismographic Network, and US/Transportable Array.

**Sustainability:** IRIS has strategic plans for its [Data Management System](http://www.iris.edu/about/DMC/DMS_Strategic_Plan.pdf) and [Education and Outreach](http://www.iris.edu/hq/files/programs/education_and_outreach/committee/mar_12/EO_BOD_120223_proposal.pdf) programs.

#### Unidata

**General Background.** Unidata is a scalable, loosely-federated model of more than 250 institutions brought together under the common goal of sharing data and tools to access and visualize data. Unidata is funded primarily by NSF and is one of eight programs in the University Corporation for Atmospheric Research (UCAR) Office of Programs (UOP). Unidata is composed of a National Center and Discipline-Specific Centers that interact directly and support their communities. Participation in the Unidata Community allows users to access data and Unidata software to explore earth-system phenomena as well as make beneficial connections through a variety of interactions.

Unidata traces its roots to a small project 25 years ago designed to provide real-time weather data to, by and large, synoptic meteorologists in U. S. universities.

**Funding/Business Model.** The Unidata receives core funding of 3.9 million per year from NSF/AGS (core funding), which covers approximately 85-90% of Unidata’s annual expenditures. This funding is renewed through a proposal process every five years. The rest of the funding comes from other grants and awards from NOAA, NASA, and other NSF divisions and programs, as Unidata does respond to solicitations.

“There is no membership fee, nor [does Unidata] charge for anything. That includes access to data, software, support, and even training... Free and open access to data and software is in the Unidata DNA, and a core value...Even the private sector doesn't pay anything, even though there are over 100 companies, large and small, that benefit from Unidata.” [[8]](#footnote-9)

**History.** “For the first 10-15 years, the entire funding for Unidata was in the single, core-funding award. As [Unidata’s] activities expanded and gained more users, more and more opportunities arose for funding from other sources. From a modest beginning of serving a few dozen meteorology programs in the late 1980s, today Unidata has users in 200+ countries, 7000 organizations, and 1500 academic institutions worldwide. Virtually every Federal Agency, including NOAA, NASA, USGS, DOD, and DOE uses Unidata's software, although Unidata does not receive funding from all of these agencies….

…Prior to 1998, NSF/ATM (back then the division was called ATM) was made every three years…Until 2008, Unidata's award from NSF was made under a larger Cooperative Agreement between UCAR [University Corporation for Atmospheric Research] and NSF…UCAR manages NCAR [National Center for Atmospheric Research], an NSF-funded FFRDC [Federally Funded Research and Development Center]. Nowadays, [Unidata] receives a separate grant from NSF GEO/AGS, and the funding is completely separate from that for NCAR.”[[9]](#footnote-10)

**Governance.** Unidata is governed by a Policy Committee and a Users Committee (standing committees), in addition to Steering Committees formed on an as-needed basis, all of which are composed of representatives from the Unidata Community. The Policy Committee “holds the primary responsibility for guiding the Unidata Program. The eight voting members of the committee are drawn from academia, balancing the membership to represent a range of colleges and universities and to reflect the concerns of both instructors and researchers… The [Policy Committee](http://www.unidata.ucar.edu/committees/polcom/) meets twice a year, and its members serve three-year terms. The committee's official charge is to make recommendations to the director of Unidata on policies, activities, and objectives. The committee's most important responsibility is to ensure that the needs of the university community are met and that Unidata is responsive to those needs.”[[10]](#footnote-11)

The Users Committee is “Composed of university representatives appointed by the [Policy Committee](http://www.unidata.ucar.edu/committees/polcom/) chair, [and] is the primary mechanism for gaining feedback on the effectiveness of the Unidata Program.” he committee is charged with three tasks:

* Determining the attitudes of the user community toward the Unidata program;
* Soliciting suggestions for additions to the data streams and software products; and
* Facilitating the exchange of ideas among users on the Unidata Program and its systems.

The [Users Committee](http://www.unidata.ucar.edu/committees/usercom/) meets a minimum of twice yearly and reports to the Policy Committee; its members are appointed for three years.”[[11]](#footnote-12)

The Unidata governance framework consists of representation by community members who are already active in their communities; tight feedback loops on program projects; direct lines of communication among the community, governing bodies, and staff; and a consensus-oriented and pragmatic approach to decision-making. Unidata Governance challenges include how to best scale efforts to meet the needs of growing communities, how to balance the need to develop new solutions with the need to provide support for existing solutions, and how to do all these things with limited resources.

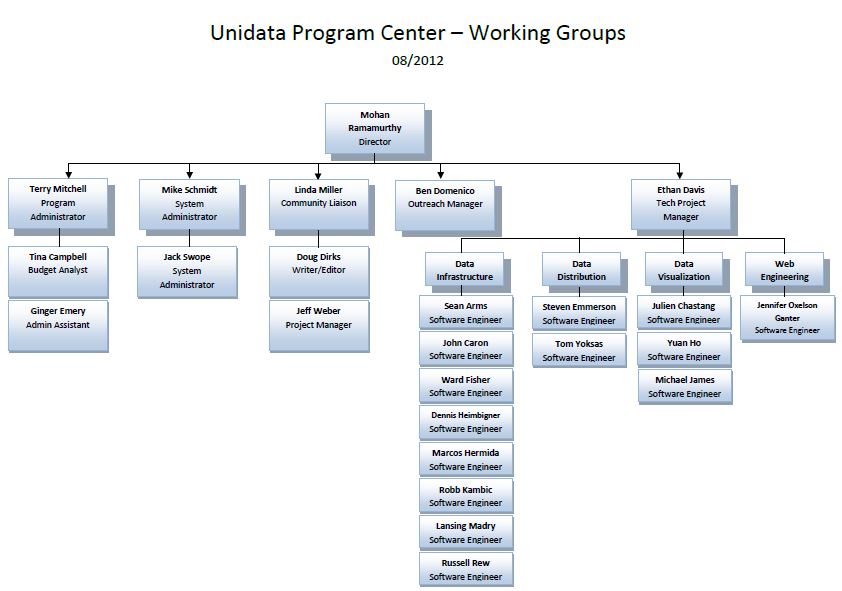


Figure . Unidata Org Chart

**Staff.** Unidata staff consists of a director and two dozen paid staff members, who are funded by the core grants from NSF and the other agencies listed in the funding section. The Unidata Governing Committee Members are all volunteers. However, the Unidata program, with funding in the core award, pays the travel expenses of Governing Committee Members to attend meetings but does not remunerate for their participation.

**Sustainability.** Unidata recently released a five year plan (<http://www.unidata.ucar.edu/publications/directorspage/Unidata_2013.pdf>) and a strategic plan (<http://www.unidata.ucar.edu/publications/2012stratplan/Strategic_plan.pdf>) with targets for the year 2020.

#### International Geo Sampler Number (IGSN)

**General Background.** SESAR (System for Earth Sample Registration) operates the registry that distributes the IGSN. “SESAR catalogs and preserves sample metadata profiles, and provides access to the sample catalog via the [Global Sample Search](http://app.geosamples.org/search.php). The IGSN or International GeoSample Number is a 9-digit alphanumeric code that is assigned to geological specimens and related sampling features such as drill holes or wells to ensure their unique identification and unambiguous referencing of data generated by the study of samples.”[[12]](#footnote-13)

“The design of SESAR and its operation are guided by the following principles:

* **Broad application**: The system should be designed in a way that it allows the broadest application to any type of Geoscience sample as requested or required. Priority should be placed on samples collected now and in the future, but the system should accommodate legacy samples.
* **Global scope**: Identifiers should be unique on a global scale because samples and sample-based data are shared on a globally.
* **Easy implementation**: The benefits to personal and institutional sample organization and data management should be maximized, and the burden minimized, in order to maximize community buy-in and appreciation. This requires development of easy and convenient sample registration procedures, including the preservation of personal sample names.
* **Sustainability**: The system needs to be maintained and operated in a professional IT environment that offers the highest level of data management standards, including access security, disaster recovery backups and procedures to ensure uninterrupted access, continuous maintenance, and long-term sustainability.
* **Maximum utility**: The system should provide valuable services to the community, and new opportunities for scientific advance that do not currently exist, which would maximize its utility and encourage its use and acceptance, and lead to appreciation of its value. Possibilities include the use of SESAR as a global sample catalog, or for management of personal or institutional sample collections.
* **“Science is the driver”**: The primary driver for its design and implementation should be the advancement of science.
* **Community involvement**: SESAR should serve the community, therefore the broad science community needs to be involved in the design, implementation, and operation of the system.”[[13]](#footnote-14)

**Governance.** Anyone can register samples with IGSN, but must first become a SESAR user. “Governance is based on the statutes there were approved at the founding assembly and that are largely determined by the German law for nonprofit organizations. [The IGSN has] an executive board with an elected president, vice-president, treasurer, and secretary. New members need to be approved by the general assembly, and [the IGSN has] just set up a date in June for the [General Assembly] to vote on new applications.  Statutes can be accessed at <http://dokuwiki.gfz-potsdam.de/datawiki/doku.php?id=igsn:start>.”

**Staff.** At this time, the IGSN office consists of three staff members who organize telecoms and the general assembly, and who operate the registry and handle service at GFZ Potsdam.

**Funding**. SESAR and IGSN are funded by the National Science Foundation. “The current business model of the IGSN eV is that each allocating agent funds its IGSN related services. Members will pay a fairly small membership fee (500 Euro at the moment) that covers these activities.” [[14]](#footnote-15) The IGSN is still in the process of setting up an account in Germany, and thus has not yet invoiced members.

#### Open Topography

**General Background:** OpenTopography “facilitates community access to high-resolution, Earth science-oriented, topography data, and related tools and resources…it is based at the [San Diego Supercomputer Center](http://www.sdsc.edu/) at the [University of California, San Diego](http://www.ucsd.edu/) and is operated in collaboration with colleagues in the [School of Earth and Space Exploration](http://sese.asu.edu/) at [Arizona State University](http://www.asu.edu/).  OpenTopography was initially developed as a proof of concept cyberinfrastructure in the Earth sciences project as part of the NSF Information and Technology Research (ITR) program-funded [Geoscience Network (GEON) project](http://www.geongrid.org/).”[[15]](#footnote-16) “GEON was designed as an equal collaboration between Information Technology (IT) and geoscience researchers, with the goal of developing an IT platform to facilitate the next generation of geoscience research and education… In addition to the computing resources provided by the project, GEON was designed to bridge cultural and disciplinary boundaries to bring together Earth and computer science experts for the common goal of developing the next generation of Earth science tools…

The GEON LIDAR Workflow (GLW) application was initially developed in 2004 as a demonstration project in a partnership between the [Active Tectonics, Quantitative Structural Geology and Geomorphology group](http://activetectonics.asu.edu/) at Arizona State University and GEON computer scientists at SDSC. Since the early proof of concept days, the GEON LiDAR Workflow system (now OpenTopography) implementation has undergone many modifications as the project has evolved from an R&D effort to a community facility…In 2009, the OpenTopography team submitted a proposal to the National Science Foundation [Earth Sciences: Instrumentation and Facilities (EAR/IF)](http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=6186) program to fund OpenTopography as a LiDAR data facility independent of GEON.  This proposal was success and OpenTopography [began operations under this new funding source in fall of 2009](http://www.opentopography.org/index.php/news/detail/nsf_awards_sdsc_arizona_state_university_17_million_for_national_opentopogr).”[[16]](#footnote-17)

Open Topography has several agreements regarding services and data hosting:

**Service Agreements:**

* [Indiana Geographic Information Council](http://www.igic.org/) - IGIS has contracted with OpenTopography to host and distribute lidar topography data collected by the [Indiana statewide ortho imagery and lidar program](http://www.igic.org/news/index.php?itemid=468).

**Memoranda of Understanding:**

* NSF [**National Center for Airborne Laser Mapping (NCALM)**](http://ncalm.org) - Through this MOU, OpenTopography will be the sole distribution pathway for all NCALM-collected lidar data.  NCALM collected data will appear in OpenTopography after the standard two year data embargo (6 months for graduate student seed proposals) provided for all NCALM supported PIs. As is the case for all OpenTopography-hosted data, NCALM datasets receive a Digital Object Identifier (DOI) to facilitate dataset access and citation. Work is also ongoing to migrate legacy NCALM data (point cloud and raster topographic data products) into OpenTopography.  OpenTopography will also provide online data archive for raw NCALM data products via the [San Diego Supercomputer Center Cloud](http://cloud.sdsc.edu).
* NSF [**Critical Zone Observatories (CZOs)**](http://criticalzone.org/) - OpenTopography and the CZOs have an MOU in place regarding hosting and distribution of CZO lidar data products.
* [**Tahoe Regional Planning Agency**](http://www.trpa.org/) - MOU with TRPA for OpenTopography hosting of the [Lake Tahoe Basin Lidar dataset](http://opentopo.sdsc.edu/gridsphere/gridsphere?gs_action=lidarDataset&cid=geonlidarframeportlet&opentopoID=OTLAS.032011.26910.1).
* [**Teton Conservation District**](http://www.tetonconservation.org/) - MOU in place with TRPA for OpenTopography hosting of the [Teton Conservation District, Wyoming lidar datatset](http://opentopo.sdsc.edu/gridsphere/gridsphere?gs_action=lidarDataset&cid=geonlidarframeportlet&opentopoID=OTLAS.062011.26912.1)
* [**Oregon Department of Geology and Mineral Industries (DOGAMI)**](http://www.oregongeology.org/sub/default.htm) - OpenTopography and DOGAMI have an MOU in place for OpenTopography to host all point cloud data collected by the [Oregon Lidar Consortium](http://www.oregongeology.org/sub/projects/olc/) (managed by DOGAMI).  This is an ongoing activity and we are progressively ingesting and releasing these data.  Available point cloud data can be found on our [DOGAMI Lidar Program page](http://opentopo.sdsc.edu/gridsphere/gridsphere?gs_action=lidarDataset&cid=geonlidarframeportlet&opentopoID=OTLAS.022011.2994.1).

**Data Partners:**

* [**EarthScope**](http://www.earthscope.org/) - Funded through a subcontract, OpenTopography was the distribution pathway for LiDAR data acquired by the [EarthScope lidar project](http://facility.unavco.org/project_support/es/geoearthscope/).  The dataset includes LiDAR for active faults throughout the western United States.  Hosting of these data is sustained under OpenTopography’s funding from NSF.
* [**Scripts Institute of Oceanography Visualization Center**](http://siovizcenter.ucsd.edu/)- OpenTopography collaborated with the SIO Viz Center to incorporate visualizations of B4 and EarthScope LiDAR topography into our portal so that they can be accessed along with the other data products.
* [**INTERFACE (INTERdisciplinary alliance for digital Field data ACquisition and Exploration)**](http://facility.unavco.org/project_support/tls/tls-interface.html) - Funded by NSF and managed by UNAVCO to help geoscientists in obtaining high resolution Terrestrial Laser Scanner (TLS) data.  We are working with INTERFACE and UNAVCO to explore TLS data hosting and processing in OpenTopography.
* [**RIT Information Products Laboratory for Emergency Response (IPLER)**](http://ipler.cis.rit.edu/) - OpenTopography is hosting the [lidar data collected following the January 12, 2010 Haiti earthquake](http://opentopo.sdsc.edu/gridsphere/gridsphere?gs_action=lidarDataset&cid=geonlidarframeportlet&opentopoID=OTLAS.072010.32618.1) in a partnership with the RIT IPLER group who oversaw the data collection.

**Cyberinfrastructure R&D:**

* [**NASA LiDAR Access System (NLAS) Project**](http://www.opentopography.org/index.php/about/nlas) - The NLAS project is a collaboration between [UNAVCO](http://www.unavco.org/), the [National Snow and Ice Data Center](http://nsidc.org/), OpenTopography, and the [NASA Goddard Space Flight Center](http://www.nasa.gov/centers/goddard/home/index.html) to develop a NASA LiDAR Access System (NLAS) to serve laser altimetry data from the NASA space-based [ICESat](http://icesat.gsfc.nasa.gov/) system and the airborne [LVIS](http://lvis.gsfc.nasa.gov) platform through OpenTopography.
* [**CyberGIS: Software Integration for Sustained Geospatial Innovation**](http://www.cigi.illinois.edu/cybergis-project/index.php) - The CyberGIS project is funded by the NSF’s [Software Infrastructure for Sustained Innovation program](http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=503489&org=NSF&sel_org=XCUT&from=fund) with the goal of establishing a fundamentally new software framework encompassing a seamless integration of cyberinfrastructure, geographical information science (GIS), and spatial analysis and modeling capabilities.  OpenTopography is one of several teams working on the project.”[[17]](#footnote-18)

**Funding:** Core operational support for OpenTopography comes from the [National Science Foundation Earth Sciences: Instrumentation and Facilities Program (EAR/IF)](http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=6186) and the [Office of Cyberinfrastructure](http://www.nsf.gov/dir/index.jsp?org=OCI).  In addition, we receive funding from the NSF and NASA to support various OpenTopography related [research and development activities](http://www.opentopography.org/index.php/about/cyberresearch). The NSF has recently renewed funding for OpenTopography. The three-year renewal under the National Science Foundation’s Geoinformatics and Earth Sciences: Instrumentation and Facilities (EAR-IF) program follows an initial three-year award from EAR-IF and the Office of Cyberinfrastructure, announced in late 2009.

**Governance**: Advisory Committee: “The seven-member OpenTopography Advisory Committee represents OpenTopography users and partners and provides valuable input on decisions related to OpenTopography priorities for data ingestion, interoperability relationships, algorithm and tool deployment, and collaborations. Members serve a three-year term with the committee chair elected from within. The committee meets one to two times per year.”**[[18]](#footnote-19)**

**Staff:** The OpenTopography team consists of 7 staff members from the [Advanced CyberInfrastructure Development (ACID) group](http://acid.sdsc.edu/) at the [San Diego Supercomputer Center](http://www.sdsc.edu/) at UCSD, [UNAVCO](http://www.unavco.org) and the [Active Tectonics, Quantitative Structural Geology and Geomorphology group](http://activetectonics.asu.edu/) at Arizona State University. Several students are also involved in OpenTopography related work.

#### OneGeology

**General Background.** OneGeology is an initiative that brings together the geological surveys of more than 115 countries, which the goal of creating public available, dynamic, geologic map data for the world. OneGeology’s participating countries cover more than 100 million square kilometers (70% of Earth’s land surface). OneGeology’s aims are to:

* “create dynamic digital geological map data for the world.
* make existing geological map data accessible in whatever digital format is available in each country. The target scale is 1:1 million but the project will be pragmatic and accept a range of scales and the best available data.
* transfer know-how to those who need it, adopting an approach that recognises that different nations have differing abilities to participate.
* [make] the initiative…truly multilateral and multinational and will be carried out under the umbrella of [several global organisations](http://www.onegeology.org/participants/organisational_bodies.html).”[[19]](#footnote-20)

“The project…involves many different stakeholders working together: the network of geological surveys around the world; the international umbrella organisations of the Commission for the Geological Map of the World (CGMW), International Union of Geological Sciences (IUGS), International Year of Planet Earth (IYPE) and United Nations Educational Scientific & Cultural Organisation (UNESCO) and the International Framework of the International Steering Committee for Global Mapping (ISCGM). It is hoped that [OneGeology] will attract other relevant bodies as the project moves forward.

The concept is a completely modern paradigm: it is planned as a distributed model – a dynamic set of geological map data served mostly on a national basis by individual geological surveys and other bodies (e.g. the polar and marine surveys and research bodies) to a web portal and as such will be frequently updated and improved by them and reflect the most up to date data they possess. To achieve its goals the project team combine[s] state-of-the-art skills in geoscience data modelling and information management with worldwide expertise and experience in lithological and stratigraphical classification. The project [is] closely interlinked with the IUGS Commission for the Management and Application of Geoscience Information (CGI) and in particular its work on a global data model and interchange standard – GeoSciML.” [[20]](#footnote-21)

**History.** “The OneGeology kick-off meeting was held in Brighton, UK in March 2007. A total of 81 geoscientists from 43 countries across the world gathered to consider a proposition – would they be prepared to collaborate to create a global geological map dataset? The answer was a resounding 'yes' and the proposition became an initiative, known as 'OneGeology'. The Brighton meeting produced a unanimous ['Accord'](http://www.onegeology.org/docs/brighton_workshop/accord/Accord_English.pdf) that provides the governance, technical and political essentials for OneGeology. Since Brighton, an international governance structure has been agreed in detail, and a [Memorandum of Understanding](http://www.onegeology.org/organisation/mou.html) has been signed by the [global bodies involved](http://www.onegeology.org/participants/organisational_bodies.html) . Brighton resulted in a large amount of interest from prospective participants, engagement and recruitment continues.

The Brighton meeting was quickly followed by a technical workshop, which was held in Utrecht, Netherlands on 30-31 May 2007. The resulting [action list](http://www.onegeology.org/docs/technical/Utrecht_technical_workshop_May07-Action_list.pdf) is available. Following Management Team discussions about Intellectual Property Rights and Data Use, a report was written, agreed and ratified by the Steering Group. The resulting [Intellectual Property Rights and Data Use Policy](http://www.onegeology.org/misc/downloads.html?Accordion1=9#IPR) is available.”[[21]](#footnote-22)

**Governance.** OneGeology is governed by a Memorandum of Understanding, signed July 2007, between UNESCO, CGMW, IUGS, IYPE, ISCGM (CGMW: Commission for the Geological Map of the World, ISCGM: International Steering Committee for Global Mapping, IUGS: International Union of Geological Sciences, IYPE: International Year of Planet Earth, UNESCO: United Nations Educational, Scientific and Cultural Organisation) and an international consortium of geological surveys. These groups form an International Coordinating Committee for Global Geoscience Mapping (ICCGGM), which meets once per year and is comprised of representatives from each of the six international organizations.

A Steering Group, comprised of representatives from the geological surveys and organizations guides and advises OneGeology. Steering Group members provide governance for the OneGeology Initiative, and represent the regions of Europe, Asia, Africa, Latin America, Northern America and Oceania. The Chair of the Steering Group also serves on the ICCGGM.

OneGeology is managed by a Secretariat, which ensures the coordination and continuity of the initiative. An Operational Management Group also provides overall coordination of OneGeology. It meets approximately once per year, and is guided by a Terms of Reference document.

OneGeology has also task groups to achieve its goals, including a Technical Working Group, which meets approximately once per year, and is guided by a Terms of Reference document. The Technical Working Group “provides a methodology and series of 'user cookbooks' to guide participants through the process.” [[22]](#footnote-23) OneGeology’s governance structure is pictured in the figure below.

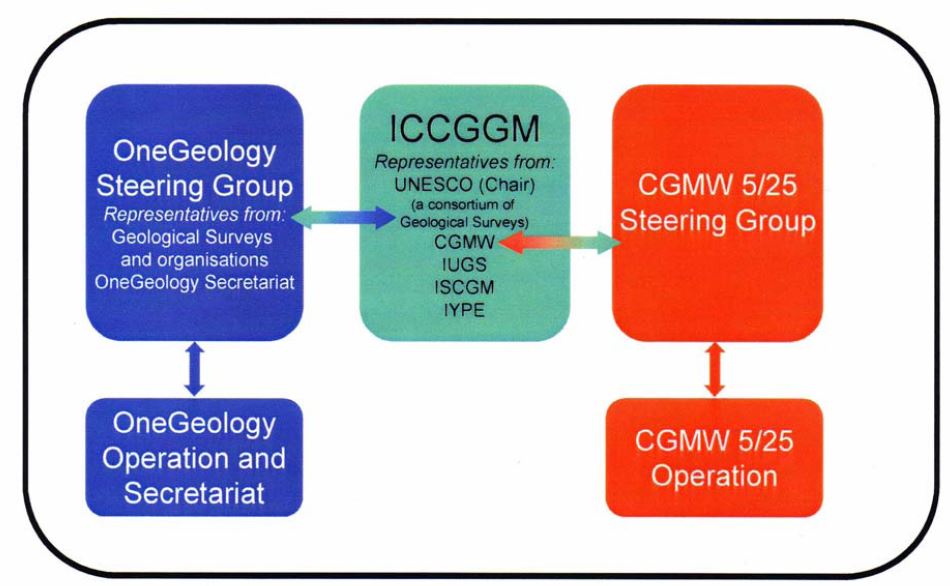


Figure : OneGeology Governance Structure[[23]](#footnote-24)

**Funding/Business Model.** OneGeology is supported by UNESCO and nine other international bodies.

**Staff.** The Secretariat is comprised of 4 staff members and is housed at the British Geological Survey.

### IT Groups

### Apache Software, DataNet, DataONE, W3C, DPN, OpenDAP, IETF, DataBasin, DataCite, Open Knowledge Framework

#### Apache Software Foundation

**General Background. “**The Apache Software Foundation, a US 501(3)(c) non-profit corporation, provides organizational, legal, and financial support for a broad range of over 140 open source software projects… The mission of the Apache Software Foundation (ASF) is to provide software for the public good…by providing services and support for many for like-minded software project communities of individuals…

The Foundation provides an established framework for intellectual property and financial contributions that simultaneously limits potential legal exposure for our project committers. Through a collaborative and meritocratic development process known as The Apache Way, Apache™ projects deliver enterprise-grade, freely available software products that attract large communities of users. The pragmatic Apache License makes it easy for all users, commercial and individual, to deploy Apache products.”[[24]](#footnote-25)

**History. “**Between 1995 and 1999, the Apache HTTPD Web Server created by the Apache Group became the leader of the market (and currently still is, with more than 65% of the web sites in the world powered by it). The foundation was created in 1999 by a group of people, that called themselves the "Apache Group" and had come together several years earlier, to continue to support and maintain the HTTPD web server written by the NCSA. But as the web grew bigger, economical interests started to grow, and the Apache web site hosted new sister projects (such as the mod\_ perl project, the PHP project, the Java Apache project). The need for a more coherent and structured organization that would shield individuals from potential legal attacks felt more and more necessary.

**Governance.** Apache governance is meritocracy-based project promotion. “Individuals who have demonstrated a commitment to collaborative open-source software development, through sustained participation and contributions within the Foundation's projects, are eligible for membership in the ASF. An individual is awarded membership after nomination and approval by a majority of the existing [ASF members](http://www.apache.org/foundation/members.html). Thus, the ASF is governed by the community it most directly serves -- the people collaborating within its projects.” [[25]](#footnote-26)

The [ASF members](http://www.apache.org/foundation/members.html) periodically elect a [Board of Directors](http://www.apache.org/foundation/board/) to manage the organizational affairs of the Foundation, as accorded by the [ASF Bylaws](http://www.apache.org/foundation/bylaws.html). The Board, in turn, appoints a number of officers to oversee the day-to-day operations of the Foundation. “The board is responsible for management and oversight of the business and affairs of the corporation in accordance with the foundation [Bylaws](http://www.apache.org/foundation/bylaws.html). This includes management of the corporate assets (funds, intellectual property, trademarks, and support equipment) and allocation of corporate resources to projects. However, technical decision-making authority regarding the content and direction of the Apache projects is assigned to each respective project management committee. The board is currently composed by nine individuals, elected between the members of the foundation. The bylaws don't specify the number of officers that the board should have, but historically, this was the number of the first board and it has never changed. The board is elected every year.”[[26]](#footnote-27)

Individual Apache projects are in turn governed directly by Project Management Committees (PMC) made up of individuals who have shown merit and leadership within those projects. There are detailed descriptions of [ASF and project governance models](http://www.apache.org/foundation/governance/).

**Funding.** The ASF Sponsorship Program “is the official avenue for substantial, non-directed monetary contributions to the ASF. It is the closest and most direct method for a corporation or individual to support the ASF. In return, ASF will make official acknowledgement and thanks for the donation via PR involvement…sponsor logo and link on the ASF Thanks page…and an official ASF logo to place on the Sponsor's site.”[[27]](#footnote-28) There are four sponsorship levels (Platinum, Gold, Silver, Bronze), defined by donation and in-return benefits of sponsorship, and each sponsorship lasts for one year, but may be renewed indefinitely.

Sponsorship helps cover the costs of over 100 top level projects, over 400 active members, over 2000 committees, countless contributors and users, millions of lines of code and the 35+ servers which make up ASF infrastructure, in addition to normal operating costs, such as bandwidth costs, servers and hardware, legal and accounting, normal office expenses, marketing and PR and sub-contractors for administrative and secretarial duties.

**Staff.** Corporate officers appointed by the Board run the ASF. Dozens of vice presidents appointed by the board serve as chairs for the PMCS of each Apache project.

**Sustainability.** The ASF has a [strategic plan](https://cwiki.apache.org/OFBADMIN/strategic-plan.html), last updated in 2010.

#### Data Observation Network for Earth (DataONE)

**General Background.** Data Observation Network for Earth (DataONE) “is the foundation of new innovative environmental science through a distributed framework and sustainable cyberinfrastructure that meets the needs of science and society for open, persistent, robust, and secure access to well-described and easily discovered Earth observational data…DataONE currently hosts three Coordinating Nodes that provide network-wide services to enhance interoperability of the Member Nodes and support indexing and replication services. Coordinating Nodes provide a replicated catalog of Member Node holdings and make it easy for scientists to discover data wherever they reside, also enabling data repositories to make their data and services more broadly available to the international community.” [[28]](#footnote-29) DataONE is also a community that provides extensive resources to its members and offers various opportunities to get involved, contribute and use data.

**Funding. “**Supported by the U.S. National Science Foundation (Grant [#OCI-0830944](http://nsf.gov/awardsearch/showAward.do?AwardNumber=0830944)) as one of the initial DataNets, DataONE will ensure the preservation, access, use and reuse of multi-scale, multi-discipline, and multi-national science data via three primary cyberinfrastucture elements and a broad education and outreach program.”[[29]](#footnote-30) **“**DataONE is a collaboration among many partner organizations, and is funded by the US National Science Foundation (NSF) under a Cooperative Agreement.”[[30]](#footnote-31)

**Governance.** DataONE is led by an Executive Team that provides overall project management, a Leadership Team (composed of Co-PIs and representatives from key institutions and focal areas), a Core Cyberinfrastructure Team, a team of Software Developers, and several post docs, in addition to an External Advisory Board that “provides advice to ensure that DataONE is fulfilling its mission, serving its user communities, implementing a sustainable business plan, and developing an evolving, strategic vision.”[[31]](#footnote-32) “DataONE engages its community of partners through [working groups](http://www.dataone.org/working_groups) focused on identifying, describing, and implementing the DataONE cyber-infrastructure, governance, and sustainability models. Additionally, the [DataONE User's Group](http://www.dataone.org/dataone-users-group) provides the opportunity for funders, users, developers, educators or any other stakeholder to gather and discuss the advancement of DataONE products and services.”[[32]](#footnote-33)

DataONE communicates its progress, activities, and events through regular newsletters, frequent new items, and via Twitter, Facebook, and LinkedIn. There is also a DataONE Users Group (Dug) for more active DataONE members. DataONE governance and management structure is depicted below (Figure 4).

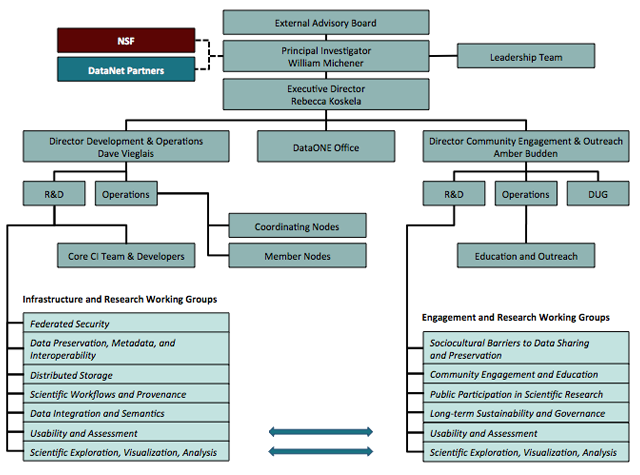


Figure 4: DataONE Organization[[33]](#footnote-34)

#### World Wide Web Consortium (W3C)

**General Background.** The W3C Consortium is an “international community where [Member organizations](http://www.w3.org/Consortium/Member/List), a full-time [staff](http://www.w3.org/People/), and the public work together to develop [Web standards](http://www.w3.org/standards/).”[[34]](#footnote-35) Its mission statement is to “lead the World Wide Web to its full potential by developing protocols and guidelines that ensure the long-term growth of the Web.”[[35]](#footnote-36) W3C is carried out through a joint agreement among three “Host Institutions” (Massachusetts Institute of Technology - MIT, European Research Consortium for Informatics and Mathematics - ERCIM, and Keio University).

W3C membership is open to any commercial, educational, governmental entity, and individuals who want to join. W3C provides very specific details on potential Member roles and has a separate set of conditions describing how W3C applies its process to membership by projects. As of May 31, 2012 W3C has 368 members.[[36]](#footnote-37) W3C’s areas of focus include architecture, technology and society, and other domains (Figure 23).

**Funding.** W3C sources of revenue include W3C Member dues, research grants and other sources of private and public funding, and sponsorship and donations. W3C fees vary depending on the annual revenues, type, and location of headquarters of an organization, while “Grants, external contracts, and other **funding outside Member dues** support a variety of work at W3C, including Member-approved [Activities](http://www.w3.org/Consortium/activities) (including Working Groups), W3C public communications, and W3C internal operations. They also help connect W3C to important communities and developments in emerging technologies and research.”[[37]](#footnote-38) W3C also has a variety of sponsorship programs to allow individuals and companies to support W3C.

**Governance.** The W3C governance framework is set up to carry out its mission statement and identifies a process for the standardization of Web technologies based on fairness, responsiveness, and progress. In administrative terms, W3C is administered by a joint agreement among several “Host Institutions” (MIT, ERCIM, Keio University, Beihang University) and “the [W3C staff](http://www.w3.org/People/) (many of whom work physically at one of these institutions) is led by a Director and CEO. A small [management team](http://www.w3.org/People/domain?domain=Management) is responsible for resource allocation and strategic planning on behalf of the staff. Regional [offices](http://www.w3.org/Consortium/Offices) play an important role in W3C being an [international organization](http://www.w3.org/Consortium/facts#international).”[[38]](#footnote-39)

Standards development is based on community consensus, and all W3C stakeholders can provide input in the development of W3C standards. Several documents, including the W3C Process Document, Member Agreement, and Patent Policy, establish the roles and responsibilities of W3C consortium members. Key organizational components include:[[39]](#footnote-40)

* **Advisory Committee:** each W3C member has a representative and participates in the W3C Process
* **Advisory Board:** members are elected by the Advisory Committee
* **Technical Architecture Group (TAG):** documents Web Architecture Principles
* **Chartered groups:** composed of Member representatives and invited experts and produces W3C deliverables

W3C fosters international participation in standards creation through a number of governance mechanisms. These mechanisms include liaisons with national, regional and international organizations; the Offices Program (promotion of W3C recommendations to developers, application builders, and other stakeholders); a policy for authorized translations of Web standards; forums around the world for people interested in standards creation; and W3C’s Internationalization Activity, which helps to make sure the web is accessible to everyone.[[40]](#footnote-41)

W3C protects the contributions of its members and developers through a Patent Policy for Working Groups. This document facilitates the development of W3C Recommendations by W3C Working Groups, promotes the implementation of Royalty-Free W3C Recommendations, and deals with patent issues that arise during and after the development of a Recommendation.[[41]](#footnote-42)

**Staff:** A Director and CEO guide staff at W3C, and a small management team and staff of technical experts are in charge of strategic planning and allocation of resources. The W3C Team includes 80 people working from locations across the globe.

#### Digital Preservation Network

**General Background.** “The Digital Preservation Network (DPN) serves as a preservation backbone for digital information of interest to the academy. DPN Depositors add digital assets to the Network by working with an individual DPN Node to ingest and preserve content. This content is then replicated to other DPN Nodes, which together form a heterogeneous network of secure, trustworthy digital archives, each operated under diverse geographical, organizational, financial, and technical regimes. Robust (bit) auditing and repair functions ensure the fixity of content over time. Intellectual property agreements ensure the succession of rights to use of the content through the Network in the event of dissolution or divestment of content by the original depositor and/or archive.

“In the DPN ecosystem, local repositories become contributing nodes, which ingest new forms of scholarship and new collections. DPN creates several federated, replicating nodes, which are digital repositories for the contributing nodes with a specific focus on long-term preservation. The replicating nodes contain redundant, dark copies of all deposits that can be brightened in cases of catastrophic loss. The diversity of the DPN nodes mitigates the risk of a single point of failure. Objects and metadata are replicated across nodes that embody organizational, technical, physical, and political diversity. A single point of failure cannot jeopardize centuries of scholarship.”[[42]](#footnote-43)

**Funding.** DPN is funded by $20,000 annual commitments from each of the member institutions for at least the first year.

**Governance.** DPN is a federation of more than 50 members (Universities, libraries and others). The Leadership Group is currently working with a small group of university presidents to develop a more formal DPN governance structure.

**Staff.** DPN is led by a Leadership and a Tech team, equaling approximately 30 staff members. These teams are distributed geographically, and meet regularly via online meetings and occasional in-person meetings.

**Sustainability.** DPN recently launched a working group focused on a developing a long-term sustainable business model for DPN. This group will focus on identifying preservation services costs and proposing a revenue model to support DPN operations.

#### **OpenDAP**

**General Background.** OPeNDAPstands for *Open-source Project for a Network Data Access Protocol*. “OPeNDAP is a framework that simplifies all aspects of scientific data networking.

[It] provides software which makes local data accessible to remote locations regardless of local storage format. OPeNDAP also provides tools for transforming existing applications into OPeNDAP clients (i.e., enabling them to remotely access OPeNDAP served data). OPeNDAP software is freely available.”[[43]](#footnote-44)

“OPeNDAP is both the name of a non-profit organization and the commonly-used name of a protocol which the OPeNDAP organization has developed. The DAP2 protocol provides a discipline-neutral means of requesting and providing data across the World Wide Web. The goal is to allow end users, whoever they may be, to access immediately whatever data they require in a form they can use, all while using applications they already possess and are familiar with. In the field of oceanography, OPeNDAP has already helped the research community make significant progress towards this end. Ultimately, it is hoped, OPeNDAP will be a fundamental component of systems which provide machine-to-machine interoperability with semantic meaning in a highly distributed environment of heterogeneous datasets. The OPeNDAP organization exists to develop, implement, and promulgate the OPeNDAP protocol. It presents the results of its work freely to the public with the hope that it will be of service in many disciplines and facilitate sharing of and access to their datastreams.”[[44]](#footnote-45)

**Funding/Business Model.** OPeNDAP is a 501(c)(3) nonprofit corporation that provides free, open source software to the scientific community. OPeNDAP is supported by the National Oceanic and Atmospheric Administration, the National Aeronautics and Space Administration, the National Science Foundation, and the Australian Bureau of Meteorology. Additionally, OPeNDAP receives open-source project support in the form of software licenses for development tools from *IntelliJIDEA IDE from JetBrains* and *YourKit Java Profiler*.

**Governance. OPeNDAP generally holds** one technical meeting a year where developers that are using the DAP and/or OPeNDAP software can meet, present their work and talk with others.

**Staff.** OPeNDAP is run by a 4-person full-time staff, divided into a team of Corporate Officers, and an Engineering Team.

#### Open Knowledge Foundation

**General Background**. “The Open Knowledge Foundation (OKF) is a non-profit organisation founded in 2004 and dedicated to promoting open data and open content in all their forms – including government data, publicly funded research and public domain cultural content. The Foundation is an international leader in its field and has extensive experience in building tools and community around open material. [Its] software development work includes some of the most innovative and widely acclaimed projects in the area. For example, [the] CKAN project is the world’s leading open source data portal platform – used by data.gov, data.gov.uk, the European Commission’s open data portal, and numerous national, regional and local portals from Austria to Brazil.” [[45]](#footnote-46)

The OKF host more than two dozen international projects, and offers a variety of services including building websites, holding training sessions on open data and code, hosting events, conducting research, and writing open source tools. “The Open Knowledge Foundation’s award winning OpenSpending project enables users to explore over 13 million government spending transactions from around the world. [It has] an active global network which includes Working Groups and Local Groups in dozens of countries – including groups, ambassadors and partners in 21 of Europe’s 27 Member States. Tens of thousands of people from a wide variety of backgrounds take part in Open Knowledge Foundation discussions and activities. [It runs] numerous events – from small hands on policy workshops and code sprints, to our annual international events, OKFestival and OKCon, the last edition of which convened over a thousand people for five days of activities in Helsinki, Finland.In 2011 the Foundation ran the Open Data Challenge, which was Europe’s biggest open data competition to date, attracting 430 entries from 24 Member States….”[[46]](#footnote-47)

“Much of [Open Knowledge Foundations’] work is done through volunteers in [an] active global network. Community activities are organized around individual [working groups](http://okfn.org/wg/), [regional communities](http://okfn.org/local/) and [projects](http://okfn.org/projects/), each focused on a different aspect of open knowledge, but united by a common set of concerns, and a [common set of traditions](http://wiki.okfn.org/Handbook/Governance) in both etiquette and process. This community is open to everyone and is organised as a meritocracy based on open discussion and tolerance. Running through all of [Open Knowledge Foundations’] activities is a strong emphasis on [decentralized collaboration](http://wiki.okfn.org/Handbook/Governance). In particular, a primary aim is to help others develop open material as well as creating it ourselves….One of [its] major functions is acting as a hub for work on open knowledge, drawing together representatives from across the knowledge society – from academics, public servants and entrepreneurs to data experts, archivists and web developers.”[[47]](#footnote-48)

**Funding/Business Model. “**The Open Knowledge Foundation Limited (OKF) is a not-for-profit organization. It is incorporated in England & Wales as a company limited by guarantee …

The Foundation is registered with the UK ICO and is on the Data Protection Register under the terms of the Data Protection Act 1998. The Open Knowledge Foundation Limited is registered for VAT in the United Kingdom…”[[48]](#footnote-49) The OKF is funded by many funders and partners, including Shuttleworth Funded, Knight Foundation, Omidyar Network, the William and Flora Hewlett Foundation, the National Endowment for Democracy, and the Alfred P. Sloan Foundation. Individuals can also support OKF via monthly or one-time donations.

**Governance.** The OKF is guided by a Memorandum and Articles of Association that details membership policies, general meetings proceedings, member votes, powers of the directors, accounts and other governance issues. The Foundation prepares regular, publicly accessible, reports that detail its activities, in addition to annual, audited accounts according to UK company standards. “Local groups and working groups are community-run, and supported by the OKF team. The team is divided into five units (Network, Knowledge, Long term projects, Services, and Operations) reflecting key focus areas of activity. Rufus Pollock (Founder and Co-Director) and Laura James (Co-Director) manage the OKF overall in terms of general management and strategic direction, and report to the Board of Directors, who are responsible for the financial and legal probity of the Open Knowledge Foundation. Further information on the Board and its activities can be found here. The Advisory Board [provides] specialist expertise and guidance.” [[49]](#footnote-50) “The Board of Directors provides the formal governance of the Open Knowledge Foundation as a whole.” [[50]](#footnote-51) Board meeting minutes are posted online and available to the public. At minimum, the seven Board members are expected to:

* Assume collective responsibility for financial and legal probity of organization
* Attendance at board meetings either via phone or in person.
* Review of all relevant materials
* Reading the board email list (generally low-traffic and limited to matters requiring board specific attention) and responding where you have specific expertise
* Participating in email list votes from time to time (roughly six of these per year)
* Shaping and overseeing the strategy and operation of the Foundation
* A general commitment to help to promote the organization and its goals

“Advisory Board members provide guidance to the Foundation as well as being consulted on key strategic issues. There are no specific obligations as an Advisory Board member. [The OKF]occasionally organize[s] conference calls for the Advisory Board which [it hopes] members will participate in…These calls are generally quite brief and provide an opportunity to consult Advisory Board members and update them on the Foundation’s activity. More often [the OKF engages] Advisory Board members individually on specific topics related to their interests, or in meetings with Rufus and Laura [OKF co-directors] to discuss general OKF strategy and other topics.” [[51]](#footnote-52)

**Staff.** More than 40 paid staff members coordinate run OKF and coordinate its activities. Staff members are distributed throughout the world, and are divided into six units:

1. Leadership Team (3 members)
2. Network Unit (10 members)
3. Knowledge Unit (7 members)
4. Long Term Projects Unit (3 members)
5. Services Unit (15 members)
6. Operations Unit (4 members)

#### DataBasin

**General Background.** Data Basin is a “science-based mapping and analysis platform that supports learning, research, and sustainable environmental stewardship.”[[52]](#footnote-53) “Data Basin is free and provides open access to thousands of scientifically-grounded, biological, physical, and socio-economic datasets. This user-friendly platform enables people with varying levels of technical expertise to:

* Explore and organize data & information
* Create custom visualizations, drawings, & analyses
* Utilize collaborative tools in groups
* Publish datasets, maps, & galleries
* Develop decision-support and custom tools

Data Basin supports researchers, natural resource managers, advocates, teachers, students, and members of the engaged public. Members create and participate in working groups where they can visualize, draw, comment, and discuss relevant topics or geographies. Data Basin breaks down barriers to collaboration and negotiation for users affiliated with universities, non-profits, tribes, companies, and local, state, federal, and national governments.” Data Basin offers services to process and upload spatial data, host spatial data, data gateways, premium mapping and analysis, Data Basin training and workshops and custom tool development.

Participants can browse the data showcase, search geographies or keywords, explore a sample map, and become a member, which connects users to “networks of spatially inspired people, expansive and scientifically-credible datasets, tools to support your exploration, customization, & communication, and educational resources & materials…Data Basin is used by interested citizens, students & educators, natural resource practitioners, and scientists from diverse sectors and geographies.” [[53]](#footnote-54) The more than 6000 members can participate in 170 groups to find collaborators and participate in topical discussions.

**Funding/Business Model.** Data Basin is run by the Conservation Biology Institute, which is a 501(c)(3) organization. The Wilburforce Foundation provides funding to maintain and develop Data Basin. Additional funding is provided by the Kresge Foundation, and individuals can also make donations to support Data Basin.

**Governance.** Terms of Use are outlined in a Disclaimer and Use Policy, posted on the Data Basin website.

**Staff.** An 11 member team develops and maintains Data Basin.

#### DataCite

**General Background.** DataCite is a not-for-profit formed in London, UK, in December 2009 by: the British Library, the Technical Information Center of Denmark, TU Delft Library, the National Research Council’s Canada Institute for Scientific and Technical Information (NRC-CISTI), California Digital Library, Purdue University, and the German National Library of Science and Technology. Since [its] foundation, [DataCite has] been joined by several more leading organisations around the world,” and now has 23 member organizations are distributed across 14 countries. DataCite offers a series of services, resources and events to support its users. The goals of DataCite are to:

* “Establish easier access to research data on the Internet
* increase acceptance of research data as legitimate, citable contributions to the scholarly record
* support data archiving that will permit results to be verified and re-purposed for future study”[[54]](#footnote-55)

DataCite “brings together the datasets community to collaboratively address the challenges of making research data visible and accessible. Members of DataCite meet in person every six months at summer and winter conferences, and collaborate in established working groups. Through collaboration, DataCite:

* support[s] researchers by helping them to find, identify, and cite research datasets with confidence
* support[s] data centres by providing persistent identifiers for datasets, workflows and standards for data publication
* support[s] journal publishers by enabling research articles to be linked to the underlying data…

…Currently, DataCite is “working with data centres to assign persistent identifiers to datasets, we are developing an infrastructure that supports simple and effective methods of data citation, discovery, and access. Citable datasets become legitimate contributions to scholarly communication, paving the way for new metrics and publication models that recognise and reward data sharing. Initially [DataCite is] leveraging the Digital Object Identifier (DOI) infrastructure, which is well-established and already widely used for identifying research articles.” [[55]](#footnote-56)

**Governance.** DataCite is “global, with member institutions offering services and advice directly where they are needed by the researchers. [It consists] of a Managing Agent, currently the German National Library of Science and Technology, with Members and Associate Members around the world. DataCite is represented by its [5] Board Members.” DataCite’s workflow relationship with other global entities is pictured in the figure below.

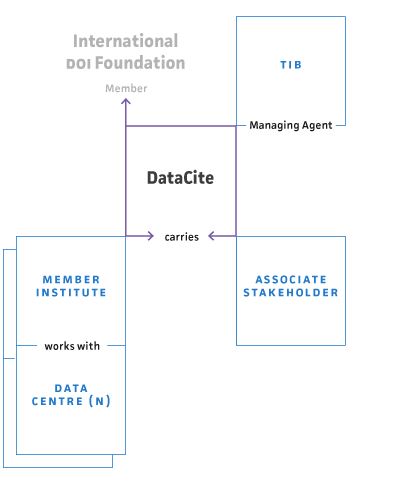


Figure . DataCite's relationship workflow and relationship with other global entities.

“DataCite has two levels of participation: full membership and associate membership. Full membership is geared towards national libraries and data centres, while associate membership is open to a broader group of organisations who support the aims and interests of DataCite.” [[56]](#footnote-57)

**Funding/Business Model.** DataCite is supported by its members. Organizations interested in membership are always welcome to apply. “Membership is open to all not for profit organisations who wish to allocate DOI names and use the Registration Agency of DataCite in their capacity as allocating agents. A member should be actively working with data centres for the purpose of issuing DOIs. The membership fee for full members is 8.500€ p.a. A member is eligible to actively take part in the working groups, has full voting rights on all decisions, and may register unlimited DOI names for themselves and their clients.” [[57]](#footnote-58)

Organizations that are interested in DataCite’s work but do not want to act as an allocating agent for DOI names can apply for an associated membership. Associated members “have an advisory function, may also actively take part in the working groups, and have the right to attend the general assembly. An associated member has restricted voting rights. The membership fee for associated members is 1.700€ p.a.” [[58]](#footnote-59)

**Staff.** DataCite is currently managed by the German National Library of Science and Technology.

#### Internet Engineering Task Force (IETF)

**General Background.** The Internet Engineering Task Force (IETF) is “is a large open international community of network designers, operators, vendors, and researchers concerned with the evolution of the Internet architecture and the smooth operation of the Internet. It is open to any interested individual.”[[59]](#footnote-60) The goal of the IEFT is to “make the Internet work better. The mission of the IETF is to produce high quality, relevant technical and engineering documents that influence the way people design, use, and manage the Internet in such a way as to make the Internet work better. These documents include protocol standards, best current practices, and informational documents of various kinds.”[[60]](#footnote-61)

The IETF works closely with World Wide Web Consortium (W3C) and International Organization for Standardization (ISO) / International Electrotechnical Commission (IEC) standards bodies.

IETF cardinal principles are outlined in its mission statement, adopted in 2004: “The IETF will pursue [its] mission in adherence to the following:

* “Open process - any interested person can participate in the work, know what is being decided, and make his or her voice heard on the issue. Part of this principle is our commitment to making our documents, our WG mailing lists, our attendance lists, and our meeting minutes publicly available on the Internet.
* Technical competence - the issues on which the IETF produces its documents are issues where the IETF has the competence needed to speak to them, and that the IETF is willing to listen to technically competent input from any source. Technical competence also means that we expect IETF output to be designed to sound network engineering principles - this is also often referred to as "engineering quality".
* Volunteer Core - our participants and our leadership are people who come to the IETF because they want to do work that furthers the IETF's mission of "making the Internet work better".
* Rough consensus and running code - We make standards based on the combined engineering judgment of our participants and our real-world experience in implementing and deploying our specifications.
* Protocol ownership - when the IETF takes ownership of a protocol or function, it accepts the responsibility for all aspects of the protocol, even though some aspects may rarely or never be seen on the Internet. Conversely, when the IETF is not responsible for a protocol or function, it does not attempt to exert control over it, even though it may at times touch or affect the Internet.” [[61]](#footnote-62)

**History.** The “first IETF meeting was on January 16, 1986, consisting of 21 U.S.-government-funded researchers. It was a continuation of the work of the earlier GADS Task Force. Initially, it met quarterly, but from 1991, it has been meeting 3 times a year. Representatives from non-governmental entities were invited starting with the fourth IETF meeting, during October of that year. Since that time all IETF meetings have been open to the public. The majority of the IETF's work is done on mailing lists, and meeting attendance is not required for contributors.

The initial meetings were very small, with fewer than 35 people in attendance at each of the first five meetings. The maximum attendance during the first 13 meetings was only 120 attendees. This occurred at the 12th meeting held during January 1989. These meetings have grown in both participation and scope a great deal since the early 1990s; it had a maximum attendance of 2,810 at the December 2000 IETF held in San Diego, CA. Attendance declined with industry restructuring during the early 2000s, and is currently around 1,200. During the early 1990s the IETF changed institutional form from an activity of the U.S. government to an independent, international activity associated with the Internet Society….The details of its operations have changed considerably as it has grown, but the basic mechanism remains publication of draft specifications, review and independent testing by participants, and republication. Interoperability is the chief test for IETF specifications becoming standards.” [[62]](#footnote-63)

**Governance.** The IEFT is “organized into a large number of working groups and informal discussion groups (BoF), each dealing with a specific topic. Each group is intended to complete work on that topic and then disband. Each working group has an appointed chairperson (or sometimes several co-chairs), along with a charter that describes its focus, and what and when it is expected to produce. It is open to all who want to participate, and holds discussions on an open mailing list or at IETF meetings, where the entry fee is currently around USD $650 per person. The mailing list consensus is the primary basis for decision of-making. There is no voting procedure, as it operates on rough consensus process.

The working groups are organized into areas by subject matter. Current areas include: Applications, General, Internet, Operations and Management, Real-time Applications and Infrastructure, Routing, Security, and Transport. Each area is overseen by an *area director* (AD), with most areas having two co-ADs. The ADs are responsible for appointing working group chairs. The area directors, together with the IETF Chair, form the Internet Engineering Steering Group (IESG), which is responsible for the overall operation of the IETF. The groups will normally be closed once the work described in its charter is finished. In some cases, the WG will instead have its charter updated to take on new tasks as appropriate.

The IETF is formally a part of the Internet Society. The IETF is overseen by the Internet Architecture Board (IAB), which oversees its external relationships, and relations with the RFC Editor. The IAB is also jointly responsible for the IETF Administrative Oversight Committee (IAOC), which oversees the IETF Administrative Support Activity (IASA), which provides logistical, etc. support for the IETF. The IAB also manages the Internet Research Task Force (IRTF), with which the IETF has a number of cross-group relations.

A committee of ten randomly chosen volunteers who participate regularly at meetings is vested with the power to appoint, reappoint, and remove members of the IESG, IAB, IASA, and the IAOC. To date, no one has been removed by a Nominating Committee, although several people have resigned their positions, requiring replacements…

…Because the IETF does not have members (nor is it an organisation *per se*), the Internet Society provides the financial and legal framework for the activities of the IETF and its sister bodies (IAB, IRTF,...). Recently the IETF has set up an IETF Trust that manages the copyrighted materials produced by the IETF.”[[63]](#footnote-64)

IETF is managed by an administrative Secretariat, based in California. A Nominating Committee reviews each open IESG, IAB and IAOC position and nominates a candidate for each one. The Nominating Committee is “comprised of at least one chair, 10 voting volunteers, 2-3 liaisons, and an advisor. The NomCom Chair is appointed between the first and second meetings of the year, and the new NomCom officially begins its work once the selected volunteers are seated following the volunteer solicitation, random selection and community review time periods.” The Nominating Committee Chair “is appointed by the Internet Society President, ensures that NomCom completes its duties on schedule and in the best interest of the community, does not vote on the selection of the candidates [and] serves as an advisor to the next NomCom.

The IESG and the IAB each provide a liaison to serve on NomCom, and ISOC Board of Trustees may also choose to provide a liaison. The ten voting members of the NomCom are selected through a complex, random process detailed in RFC 3797 (Publicly Verifiable Nominations Committee (NomCom) Random Selection). In order to serve on the NomCom, an individual must have attended at least 3 out of the last 5 IETF meetings. Once the NomCom has recommended candidates for the open positions, the candidates are reviewed and confirmed by another body. The IAB is responsible for confirming IESG candidates, the ISOC Board of Trustees confirms candidates for the IAB and the IESG confirms the candidate for the IAOC position.” [[64]](#footnote-65)

**Funding/Business Model.** The IETF generates funds by charging meeting participation fees, and receives funds from meeting sponsors and the Internet Society through its organizational membership and the proceeds of the Public Interest Registry.

**Staff.** The Administrative Director (IAD) is responsible for day-to-day fiscal and administrative support to IETF through other activities, contractors and volunteers. All other participants and managers are volunteers, though their work is usually funded by their employers or sponsors.

### Federated Groups: ESIP, OGC

#### Federation of Earth Science Information Partners (ESIP)

**Background information.** “The Federation of Earth Science Information Partners (ESIP Federation) is a broad-based community drawn from agencies and individuals who collectively provide end-to-end handling for Earth and environmental science data and information. The ESIP Federation was founded in 1998 by NASA in response to a National Research Council (NRC) review of the Earth Observation System Data and Information System (EOSDIS).

Beginning with 24 NASA-funded partners, the ESIP Federation's purpose was to experiment with and evolve methods to make Earth science data easy to preserve, locate, access and use by a broad community encompassing research, education, and commercial interests. The ESIP Federation began as a distributed organization that is linked primarily by the Internet (i.e., a virtual organization), continuing successfully to this day to provide an evolving mechanism by which the community could voluntarily come together and act to define and serve their collective best interests.

By 2001, the ESIP Federation created a nonprofit corporation called the Foundation for Earth Science (Foundation). Through a Memorandum of Understanding with the ESIP Federation, the Foundation provided management support to the ESIP Federation as it moved from an operational prototype to an independent organization. In 2002, Foundation staff was hired to support the work of the ESIP Federation. The Foundation helped create operating policies for the ESIP Federation and facilitated the development of its first strategic plan, adopted by the ESIP Federation’s Assembly in 2004. The 2004 strategic plan reflected the evolving role that the ESIP Federation sought to play in the ensuing years, reflecting the broadening of the ESIP Federation’s base from its original core to its then 75 partners. The 2004 Strategic Plan conveyed a vision that served as an important rallying point around which the ESIP Federation sharpened its focus and continued its growth.”[[65]](#footnote-66)

**Funding.** The ESIP Federation is supported by NASA and NOAA.

**Governance Structure.** The ESIP Federation is governed by its Assembly, the standing body which affords each partner one vote. Governance documents include a Constitution and Bylaws as well as an updated 5-year strategic plan (<http://wiki.esipfed.org/index.php/Federation_Documents>) . In between annual Assembly business meetings, the ESIP Federation is governed by its Executive Committee; a body comprised of the ESIP Federation’s elected President and Vice President, its Standing and Administrative Committee chairs, and members representing each ESIP type (Data Center, Research, and Application Developer). All Executive Committee members are either elected by the Assembly or by their ESIP Type caucus. While the ESIP Federation remains unincorporated, the Foundation for Earth Science (a 501(c)3 provides financial and other management services to the ESIP Federation.[[66]](#footnote-67)  
  
**Staff.** The ESIP Federation runs with support from two professional staff and a host of volunteer contributions from experts across the earth science data and technology community.

**Sustainability.** “As the ESIP Federation entered its second decade, it became obvious that, as a consequence of its continued evolution, it was time to revisit the strategic planning process. Beginning in July 2007 in Madison, Wisconsin, a Strategic Planning Working Group was formed to develop a new vision of the ESIP Federation in its second decade. This decade will capitalize on the growth of the ESIP Federation and its concomitant diversification of its funding.

The 2008 strategic plan celebrates the special status the ESIP Federation occupies as a 10-year old organization, having employed a variety of tools to facilitate communication and interaction using both traditional and virtual tools. The ESIP Federation’s strength continues to come from its more than depth of partner organizations, including all NOAA, NASA and USGS Earth observing data centers, government research laboratories, research universities, modelers, education resource providers, technology developers, nonprofits and commercial enterprises. The organization will rely on its internal communities to set priorities for the implementation of the ESIP Federation’s strategic vision.”[[67]](#footnote-68)

### Open Geospatial Consortium (OGC)

**General Background.** OGC is a Voluntary Consensus Standards Organization and is an international industry consortium of 475 companies, government agencies and universities participating in a consensus process to develop publicly available interface standards. OGC Standards support interoperable solutions that "geo-enable" the Web, wireless and location-based services and mainstream IT. The standards empower technology developers to make complex spatial information and services accessible and useful with all kinds of applications.”[[68]](#footnote-69) OGC has close to 40 adopted standards, hundreds of product implementations, broad user community implementation worldwide, and has alliance partnerships with 30+ standards and professional organizations. OGC is member-driven. OGC’s Approach for Advancing Interoperability includes the Interoperability Program (IP), Standards Program, Compliance Program, and Marketing and Communications Program.

“OGC was founded with eight charter members at the time of its first Board of Directors meeting on September 25, 1994. From 1994 to 2004, the [membership](http://www.opengeospatial.org/ogc/members) has grown from 20 to more than 250 government, academic, and private sector organizations…In 2000 OGC founded The Open Geospatial Consortium (Europe) Limited (OGCE). OGC Austral-Asia, also known as OGC-A, followed in 2003. A third related organization, the OGC-Interoperability Institute (OGCII), was organized in 2004…”[[69]](#footnote-70)

**Funding.** OGC funding comes from membership dues and project fees (Interoperability Program initiatives).[[70]](#footnote-71)

**Governance.** OGC’s Interoperability Program, Standards Program, Compliance & Interoperability Testing & Evaluation (CITE) Program, and Marketing and Communications are governed by a Board of Directors, Strategic Member Advisory Committee, Global Advisory Council, and Executive Director (see figure 5 below). Each program then has its own governance framework and mechanisms designed to carry out that program’s goals. For example, “In the OGC Standards Program the Technical Committee and Planning Committee work in a formal consensus process to arrive at approved (or "adopted") OGC® standards,”[[71]](#footnote-72) while the “Interoperability Program is a series of hands-on engineering initiatives to accelerate the development and acceptance of OGC standards.” [[72]](#footnote-73)

The basic model for standards development is for a group of OGC members to submit a proposal to start a Standards Working Group for the specific technology and version (e.g., OGC Sensor Observation Service version 2.0). The group forms, subject to review by the OGC Architecture Board and a vote by the OGC members. The SWG does its work, and when they are ready (a few months to a few years later) the completed specification documents are submitted for adoption vote by the OGC members.

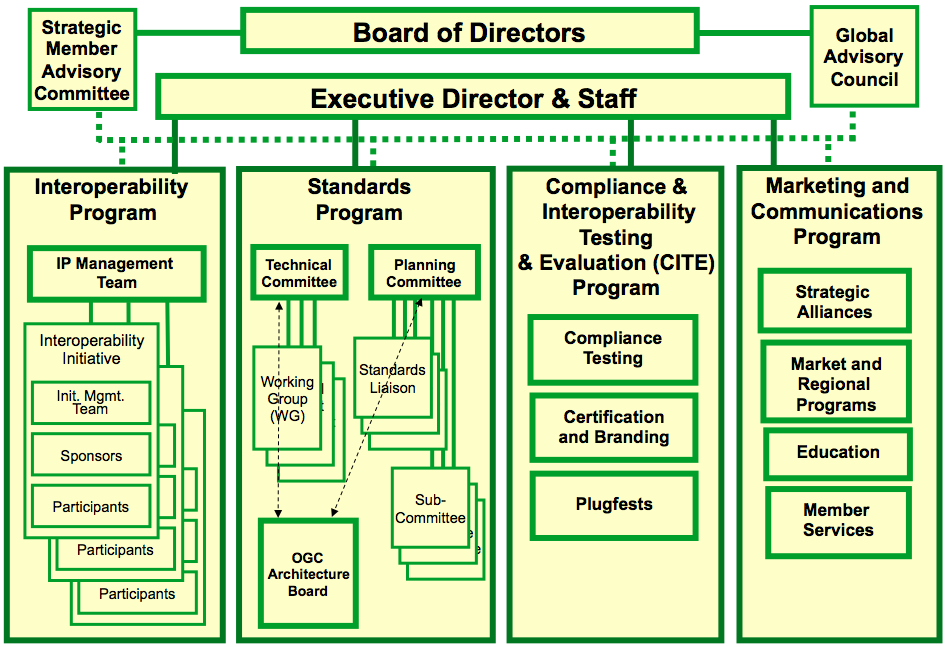


Figure 5: OGC Organization and Programs

OGC has a set of [Policies and Procedures](http://www.opengeospatial.org/ogc/policies) to fulfill its [mission and vision](http://www.opengeospatial.org/ogc/vision):[[73]](#footnote-74)

1. Current Bylaws
2. Principles of Conduct
3. Technical Committee Policies and Procedures document for Standards Development.
4. Legal Notices & Term and Conditions
5. Intellectual Property Rights (IPR)
6. Interoperability Program Policies, Processes, and Procedures (IP PP&P)
7. Policies & Procedures for Translation of OGC Public Documents
8. Policies & Procedures for the OGC Architecture Board (OAB)
9. Policies & Procedures for the Compliance Testing Program (CITE)

**Staff.** "Director, Interoperability Programs" are the program & project managers for Interoperability Program initiatives. OGC staff cannot be working group or subcommittee chairs.  At times OGC hires a consultant to lead an interoperability initiative, depending on resourcing and expertise available. These are drawn from a pre-qualified group called the IP Pool (<http://www.opengeospatial.org/projects/ippool>). Some are contributed by member companies; some are paid by project funds.[[74]](#footnote-75)

# Appendix II: USGIN Protocols, Specifications, and Community Standards

The following table represents a list of community standards, profiles, protocols, and software employed by USGIN to address the challenges associated with sharing large amounts of diverse data stored in geographically disparate locations. In general, these are not the only solutions available, but indicate resources that the developers have found useful.

| Category: [Applications](http://usgin.org/specifications/applications) | | |
| --- | --- | --- |
| [Specification](http://usgin.org/specifications?&order=field_specification_label&sort=asc) | [Description](http://usgin.org/specifications?&order=field_specification_description&sort=asc) | Uses |
| [Amazon Web Services](http://aws.amazon.com/) | Computing infrastructure resources available for rent. [More…](http://lab.usgin.org/applications/amazon-web-services) | Several servers that host USGIN Websites or provide USGIN services are provided by Amazon Web Services. |
| [Apache Tomcat](http://tomcat.apache.org/) | Free-and-open-source servlet engine. Many key applications for providing Web services can be hosted using Tomcat. [More...](http://lab.usgin.org/applications/apache-tomcat) | Used to host software applications, such as GeoServer and GeoPortal, which provide [Open Geospatial Consortium (OGC)](http://www.opengeospatial.org/)-compliant [Web services](http://usgin.org/glossary#web_service) |
| [Django](http://www.djangoproject.com/) | Open-source Web application framework written in the Python programming language. [More...](http://lab.usgin.org/applications/django) | Used to develop a number of USGIN Web applications, including the USGIN URI redirection engine. [More…](http://lab.usgin.org/groups/using-django-usgin) |
| [Drupal](http://drupal.org/) | Modular, extensible, actively maintained free-and-open-source Web site content management framework; capable of running on any PHP-capable Web server. [More...](http://lab.usgin.org/applications/drupal) | Websites related to the USGIN Initiative built using the Drupal include:  The [AASG Geothermal Data](http://www.stategeothermaldata.org/) Website  The [NDGS Portal](http://geothermaldata.org/), USGIN Lab site  The [USGIN](http://usgin.org/) Website. [More…](http://lab.usgin.org/groups/drupal-development) |
| [ESRI ArcGIS](http://esri.com/) | Commercial geographic information system software and Web service-compatible server software. Supports both proprietary service protocols ('Geoservices API'), and OGC services. | Commonly used for deploying geospatial data as [Web services](http://usgin.org/glossary#web_service). |
| [GeoServer](http://geoserver.org/) | Free-and-open-source Web service-compatible server software. [More...](http://lab.usgin.org/applications/geoserver) | Southern Methodist University, a contributor to the [National Geothermal Data System](http://geothermaldata.org/), uses GeoServer to host Web services; the landing page is [here](http://geothermal.smu.edu/geoserver/web/). |
| [PostGIS](http://postgis.refractions.net/) | Extension for PostgreSQL relational database to support geographic objects and operations. [More...](http://lab.usgin.org/applications/postgresql-and-postgis) | PostGIS databases are used to persist spatial data for ArcGIS, GeoServer, Django, and many other applications. |
| [PostgreSQL](http://www.postgresql.org/) | Free-and-open-source object-relational database system with over 15 years of active development. [More...](http://lab.usgin.org/applications/postgresql-and-postgis) | PostgreSQL is used by GeoPortal, GeoNetwork, CKAN. [CKAN is used by data.gov] |
| [Python](http://www.python.org/) | A programming language. [More...](http://lab.usgin.org/applications/python) | Python is used for USGIN Django sites and converting Excel spreadsheets into ISO metadata (the application responsible for which can be found at <http://github.com/usgin/csvtometadata>). |

# Appendix III: Recommendations for the Future of the U.S. Geoscience Information Network

## Introduction

The U.S. Geoscience Information Network (USGIN) is a partnership of the Association of American State Geologists (AASG) and the U.S. Geological Survey (USGS), who formally agreed in 2007 to develop a national geoscience information framework that is distributed, interoperable, uses open source standards and common protocols, respects and acknowledges data ownership, fosters communities of practice to grow, and develops new Web services and clients. The National Science Foundation (NSF), the Department of Energy through a connection with the National Geothermal Data System, and the USGS through a partnership with the ScienceBase project jointly funds USGIN.

This document provides a recommended road map and objectives for USGIN activities over the next 5 years, based on discussions and input from a working group consisting of USGIN stakeholders (Table 1.) The Working Group held several teleconferences over an 8 week period and focused on two interrelated topics: (1) how community-driven programs approach the issue of sustainability and (2) development of a strategic direction for USGIN to advance its goals and objectives.

**Table 1. Working Group Participants**

## 

|  |  |
| --- | --- |
| **Name** | **Affiliation** |
| David Arctur | Open Geospatial Consortium (OGC) |
| Bob Cook | Oak Ridge National Laboratory (ORNL), DataOne |
| Rob Fatland | Microsoft |
| David Ferderer | US Geological Survey (USGS) |
| Ted Habermann | National Oceanic and Atmospheric Administration (NOAA) |
| Viv Hutchison | US Geological Survey (USGS) |
| Carol Meyer | Earth Science Information Partners (ESIP) |
| Anna Milan | National Oceanic and Atmospheric Administration (NOAA) |
| Satish Sankaran | ESRI |
| Stephen Richard | Arizona Geological Survey, USGIN |
| Erin Robinson | Earth Science Information Partners (ESIP) |
| Jerry Weisenfluh | Kentucky Geological Survey |
| Ilya Zaslavski | Consortium of Universities for the Advancement of Hydrologic Science (CUAHSI) |

## Background

Representatives of the Association of American State Geologists (AASG) and the U.S. Geological Survey (USGS) met in Denver February 21-22, 2007, to discuss opportunities for making their data more accessible and interoperable across agencies. They recommended that the USGS and State Geological Surveys work together to create a distributed, national “Geological Information Network” (GIN) of digital Earth Science data using common standards and protocols, preserving ownership, credit, and control of data, and building on existing data systems (AZGS Open-file Report 2008-01, 2008).

The intention of the US Geoscience Information Network is to benefit the geological surveys by reducing the cost of online data publication and access provision, and to benefit society through easier (lower cost) access to public domain geoscience data. This information supports environmental planning, resource-development, hazard mitigation design, and decision-making. GIN supposes that sharing resources for system development and maintenance, standardizing data discovery and creating better access mechanisms, causes cost of data access and maintenance to be reduced (see Shapiro, 2000). A study by the German Institute for Standardization concluded that the economic benefits of standardization range between 0.2 and 0.9% of the gross national product (DIN, 2000; Blind et al., 2011). These studies focused on standardization in a wide variety of business domains, we suggest that they also apply in the informatics domain. Although anecdotal, consider how the music industry landscape has changed with standardized file formats and metadata schemes for recordings, or the seamless connection of most printers to computers using standard interfaces and interchange formats. Standardized access to rich data resources will create collaborative opportunities in science and business. Development and use of shared protocols and interchange formats for data publication will create a market for user applications, facilitating geoscience data discovery and utility for the benefit of society.

Since the 2007 meeting, the geoinformatics community has continued to evolve, with the emergence of new activities and partnerships that may impact USGIN. Following are some examples:

The National Geoinformatics Community (NGC) is an informal collection of academic institutions and projects that advance geoinformatics at all levels via outreach, advocacy, and fostering communities of practice. The NGC concept has evolved from several workshops, town hall meetings, and the work of an exploratory committee that met with existing and successful community efforts such as UNAVCO1, Incorporated Research Institutions for Seismology (IRIS)2, Joint Oceanographic Institutions (JOI)3, and Consortium of Universities for the Advancement of Hydrologic Science, Inc (CUASHI)4. Identified objectives for NGC include cataloging and communicating community efforts, identifying and facilitating development of community standards/specifications, and gaining a better understanding of the geoinformatics-related requirements of geoscientists and educators. No formal organizational structure has been established at this point for NGC.

The USGS Community for Data Integration (CDI) was established in 2009 to provide a forum within the USGS to facilitate data discovery, exchange, and interoperability of data and information for scientific computing, to create awareness of relevant data products, and to promote interdisciplinary science through the interaction and integration of scientific information. The CDI also provides a mechanism to deploy

1 http://www.unavco.org/unavco.html 2 http://www.iris.edu/hq/ 3 http://www.oceanleadership.org/2004/joint-oceanographic-institutions-to-lead-us-efforts-in-iodp/ 4 http://www.cuahsi.org/

consistent processes, protocols, and data management to implement the USGS Data Integration Strategy5. The CDI is supported by the Core Science Systems (CSS) Mission Area of the USGS in partnership with the USGS Office of Science Quality and Integrity. Community-led working groups meet monthly, and annual workshops encourage collaboration, showcase and exchange ideas, and create networking opportunities across the USGS partner network. Additionally, CSS supports one staff member to host a virtual monthly meeting, keeping the community informed of innovative data integration contributions occurring in the USGS and by USGS partners.

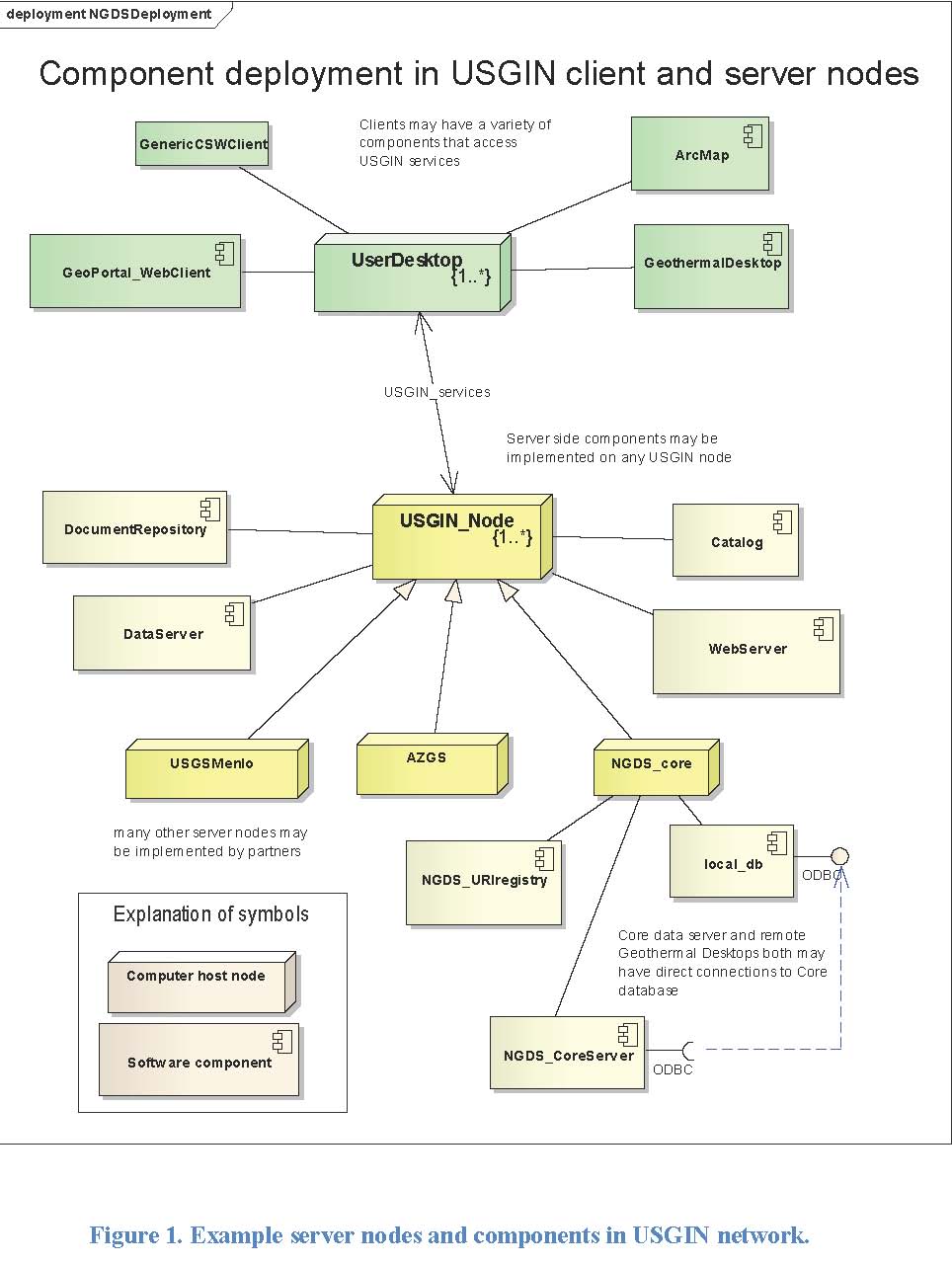
• The Federation of Earth Science Information Partners (ESIP) is a broad-based, distributed community of data and information technology practitioners that voluntarily come together to leverage collaborative synergies that create coordinated interoperability efforts across domain-specific communities. Participation in the ESIP Federation allows members to expose, gather, and enhance their own in-house capabilities in support of their organization’s own mandates. The ESIP Federation has a 13-year track record of working on Earth science interoperability solutions that connect distributed and heterogeneous communities. Ultimately, these collaborations allow data interoperability a greater chance of success. The ESIP Federation’s operations are managed by the Foundation for Earth Science and governed by a slate of officers and several standing committees, and two staff members. Development efforts are community driven, through the formation of working groups and clusters.

These activities and others may provide further opportunity for USGIN to develop a sustainability model that includes key partners by broadening the focus and leveraging activities under way.

## USGIN Today

The basic hardware and software components of the GIN system are diagrammed in Figure 1. In the past few years, USGIN has progressed from a conceptual idea to deployment of tools and capabilities critical to efficient access and dissemination of geosciences information. Examples include:

* A metadata catalog and search tool using the ESRI Geoportal, a free, open-source software product6.
* An ISO metadata profile7 that uses the OGC Catalog Service for the Web (CSW 2.0.2) as the service protocol.
* Formal collaboration agreements with GEON and the USGS National Digital Catalog to integrate their various catalogs resources. Catalog service specifications have been tested on the ESRI Geoportal tool and implemented with Deegree and Geonetwork.
* A USGIN file repository and Web tool to create metadata, register resources, and upload files. This is deployed as a Drupal v6 application8. The application pushes metadata to a Web-accessible directory for harvest into the Geoportal-based catalog when resources are published.
* A collection of Geography Markup Language (GML) simple-feature content models for use by State Geological Surveys to publish data to the National Geothermal Data System (NGDS)9. These models are built as Microsoft Excel spreadsheets to facilitate data loading; deployed services use an XML encoding of the content model as an interchange format.
* Open Geospatial Consortium (OGC)10 Web Feature Services (WFS), deployed for the NGDS by the AZGS, the Illinois Geological Survey, and Kentucky Geological Survey. GML simple feature models are used because ArcGIS and various open-source software packages can consume WFS serving simple features. Thus, a widely deployed client platform is already available for service utilization.
* A Uniform Resource Identifier (URI) redirection application to enable linked data architecture by using http URIs11. The USGIN URI Redirection Engine is considered the backbone for USGIN’s evolving linked data system. URI’s directed at the URL http://resources.usgin.org/uri-gin/ can be rewritten using rules based on regular expressions. This application is used to resolve feature identifiers to documents that describe the feature.



5 http://www.columbia.edu/~rb2568/rdlm/Faundeen\_USGS\_RDLM2011.pdf 6 http://catalog.usgin.org/geoportal 7 http://lab.usgin.org/usgin-iso-metadata-v1-1 8 http://repository.usgin.org/ 9 http://www.geothermaldata.org/ 10 http://www.opengeospatial.org/11 http://resources.usgin.org/uri-gin/uri-description/ 4

Additionally, current USGIN team members from the AzGS have been participating in various standards projects. Recent activity includes:

* Contributions to the international GeoSciML markup language 12
* Participation in the Energistics Metadata Work Group13 that is contributing to revision of the ISO 19115 and 19119 metadata standards 14
* Development of vocabularies for rock type and numerous geologic feature properties for use in populating GeoSciML instance documents 15
* AASG-USGS development of a relational data base format for geologic map data 16

USGIN community activities are documented on a Web site (http://lab.usgin.org) that includes reference information and links for services, specifications, and applications that are in use or are being considered for network use. Development activities are documented in blogs. USGIN invites anyone interested in contributing to the effort to register and contribute content to the Web site.

## Achieving Sustainability: A Five-Year Vision

The USGIN Working Group envisions further development of tools and capabilities, in addition to extending the community of practice that currently involves geoinformatics practitioners from the USGS and State Geological Surveys. Promoting engagement and participation of the state geological surveys, and increasing communication between the states, USGS, and other stakeholders are prerequisites for community development. A key element of community building is personal interaction; face-to-face meetings take time and money. We propose that maximum impact can be achieved by using the existing USGS CDI, Open Geospatial Consortium (OGC), and ESIP meetings to bring stakeholders together17.

Within this framework, the USGIN community can establish an identity for geological survey informatics practitioners, can assist in prioritizing technical development that is specific to the geological survey community, and can leverage development taking place in the larger community. Policies, protocols, and procedures for developing, reviewing, and distributing specifications can be adopted from established practices developed by existing organizations, for example the OGC. Documenting and promoting best practices through demonstrations, education, and outreach within the geological survey community is paramount for fostering deployment of interoperable services for data discovery and distribution.

These presuppositions and objectives predicate priorities for the next five years:

• Community building

o Promote face-to-face engagement with stakeholders by supporting participation in CDI, OGC, and ESIP meetings (immediate)

12 http://geosciml.org 13 http://www.energistics.org/metadata-work-group 14 http://www.energistics.org/metadata-work-group 15 https://www.seegrid.csiro.au/wiki/CGIModel/ConceptDefinitionsTG 16 NCGMP09, http://ngmdb.usgs.gov/Info/standards/NCGMP09/ 17 The GSA and AGU national meetings are such large and diverse assemblies that it is difficult to achieve the small group personal interaction necessary to foster community development. AASG annual meetings provide an excellent forum for education and outreach to the state geologists, and obtaining feedback on priorities.

o Organize coordinating committee to shepherd community

• Prioritize effort

o Nucleate efforts based on program and project requirements and personal interests

o Identify specific deliverable products (two test beds, 6-12 months; ongoing for duration)

• Improved communication

o Foster online collaboration in groups with particular objectives

• Deliver products

o Demonstrate capabilities and usefulness (18-24 months; 6-12 months after deliverables are identified)

Develop and disseminate outreach and educational materials

o Workshops, tutorials, online resources, publications (12-36 months, ongoing)

Although these objectives initially are sequential, as the community evolves all of these will need to proceed in tandem. Approximate time horizons are indicated for key steps in the process for some initial high priority activities.

A critical component to help achieve the vision for a Geoscience Information Network is to reinforce the development of a community of practitioners. To foster a sense of identity and organization for the community, we recommend formation of a coordination group with representatives from the scientific and IT communities. This group will consist of representatives from the USGS, State Geological Surveys, and the broader community. It should be small enough to be agile, comprising 7 members, with 2 or 3 representatives from the states, USGS, and large geosciences communities.

Community development is beginning to occur through collaborations within the CDI at the USGS, and through the AASG Geothermal Data project managed by the Arizona Geological Survey. Recruitment and training to bring in individuals interested in the nexus of information engineering and geoscience should remain an ongoing priority for GIN. The Working Group proposes that growth of the community should be reinforced by collaborating on two test bed activities (suggestions outlined below), and engaging with more experienced communities at the Open Geospatial Consortium and ESIP. Depending on how the priorities are established, the test bed efforts will test and develop best practices, data publication specifications, and interoperability formats using map, feature, and observation services. Data registration, catalog, and discovery specifications should be enhanced to promote accessibility. Activity organized around specific priorities and objectives is essential so that participants receive a return on their investment in time and effort, and see tangible progress. Success in the OGC community is testament that test bed projects have fostered communication, alignment of activities, and exchange of expertise and capabilities in that community.

Historically, a small number of geoinformatics practitioners have been spread across a wide variety of discipline-focused organizations, effectively diluting the knowledge base. The number of practitioners is growing, and is reaching a point where a critical mass of individuals can be brought together to meld into a community. Identifying one or two annual meetings as gathering points for this community can foster this newly forming coalescence. This Working Group recommends the USGS CDI meetings, OGC quarterly meetings, and ESIP meetings as obvious candidates for consideration.

## Implementation

Evolution of the current Balkanized geoinformatics practice into a more cohesive and effective community has been and will continue to be an incremental process. The role of USGIN as an entity in this larger community requires organization, planning, promotion, and funding. Additionally, as a member of a community activity, the role of USGIN as a leader in the community must be organic and emergent. However, there are some implementation activities the Working Group identifies as essential for USGIN, listed as recommendations below.

## Establish a Long-Term Governance Model

If USGIN is to represent the interests of the geological surveys in the larger geoinformatics community, a strong governance model is necessary to define USGIN as an entity. Such a model should contain information about the USGIN member base, designated speakers for the organization, the source of their authority, and how decisions for priorities and recommendations are made. A coordination group can be the starting point for a formal organization; one of its first major activities, however, will need to be a formal charter that defines the governance model ­membership, leadership selection and terms, staffing, and decision making processes for the organization.

## Develop a Business Model

Depending on decisions about the activities and responsibilities of a USGIN organization established in a governance charter, a defined level of financial support will be necessary and should be detailed in a business plan. Based on a cursory survey of existing organizations with intent similar to USGIN conducted by the Working Group, we recommend that a non-profit foundation model may be the most appropriate. This would define USGIN as a legal entity that could enter into contracts, receive funding, pay salary, and make grants of funding. A formal legal entity could support a small number of full time employees who look after day-to-day operations, planning, and logistics for community events. An important issue to be determined by the governance authority is whether there should be a stand-alone organization, or if it is more effective to form a group within the framework of ESIP, OGC, or some other existing organization. The implications for a separate business entity or affiliation with an existing organization need to be explored in the context of financial and intellectual property considerations that may result from the government agency nature of most geological surveys.

The business structure to support a community geoinformatics development process should promote grass root nucleation of interest groups that are the core of development efforts. Selection of groups to be nurtured by available funding should be based on community assessment of priorities and user demand. Resource allocation by government agencies in support of these activities must be tempered by programmatic priorities. Loosely knit organizations such as ESIP and the USGS-CDI provide organizational models, and the USGIN community will progress most efficiently by utilizing these existing groups to gather input and effort from the communities they attract.

Because many geological surveys have data archive and dissemination functions as part of their portfolio, some support for USGIN might be built into their operating expenses and overhead as users come to expect access to data resources through standardized network interfaces. Sharing of resources and reuse of components through USGIN will reduce the cost of these activities. Wide adoption of similar software, protocols and practices increases the number of stakeholders with an interest in supporting USGIN. Services and data streams deployed by the geological surveys must have sufficient value to the user community to support either a pay-for-use model, or to motivate continued public funding if the system it to be viable in the long term. USGIN can be seen as a metaphor—it is building an online version of the bricks and mortar libraries that historically have been the anchor of knowledge preservation and access for the geosciences.

## Explore Testbed Opportunities

A testbed is a platform for technological experimentation in a framework that allows for rigorous, transparent, and replicable testing of new technology (http://en.wikipedia.org/wiki/Testbed; Percivall, 2006). The Working Group agreed that one or two testbed activities should be proposed to engage geological surveys in development, adoption, and deployment of standard interchange protocols and document formats. This series of activities represents a good starting point to nucleate USGIN. Several test bed activities were suggested:

OneGeology – US testbed. The USGS and AASG collaborate to implement WMS and WFS services according to the OneGeology profile (GeoSciML portrayal for WMS, GeoSciML WFS for supporting WFS). USGS handles national maps (surficial deposits, glacial deposits, bedrock geology of US, etc.) and State Geological Surveys deploy services for state-scale geologic maps. The community must agree on an integrated portrayal scheme for services to attain some interoperability between map services.

Observation service test bed— the state geological surveys and USGS deploy observation services for geochronologic data, geochemical data, gravity stations, water quality data, or some similar commonly available site or sample-based data. The community would need to agree on service profiles and interchange formats, as well as procedures to avoid data duplication. This testbed activity could build on and enhance both the National Geothermal Data System and EarthChem.

Integrated catalog capability – the achievement of interoperable metadata – demonstrate an ability to harvest between catalog nodes hosted by state surveys and USGS, execute search against multiple nodes, edit tools that work against multiple nodes (with user authentication and access control).

Each of these proposed testbeds would utilize existing service protocols and interchange formats to the maximum extent possible, as well as off-the-shelf-open source software or widely deployed commercial software (e.g. ArcGIS) for service deployment. Service profiles would also consider existing clients or client development frameworks (OpenLayers, Flash, etc.) in their design.

The protocols and interchange formats used for testbed activities will be developed with consideration of existing technology and standards (following the founding principles of USGIN). The test bed activities in the geological survey community will define a scope and provide a foundation to promote the use of specifications developed in our community by the larger geoinformatics community. Adoption of some of these specifications as ‘standards’ by USGS and AASG for use by those organizations will lend authority and motivate wider adoption.

The transition from use case and test bed activities to production deployments and agreement on ‘standard’ specifications for data discovery and access must be propelled by active interest from the user communities who have a stake in the outcome. Part of the testbed planning should include identifying and contacting target communities, and exploring possibilities that they might contribute to the costs of system development and maintenance.

## Develop Marketing Strategy

Development of a USGIN education and outreach strategy to inform and engage data providers of GIN direction, infrastructure, and activities is recommended for USGIN’s development. Stakeholders will need to know how to realign their existing approaches to data delivery to best interact with USGIN, what new resources are available, how to use them, and any new opportunities for information utilization. A marketing plan should also include monitoring of network resource usage, and collection of input from the user community to identify what is working well, what needs fixing, what is not being used, and what new capabilities would be useful.

Development of educational material, giving talks and running workshops requires significant time and effort, and is generally difficult to support with project-based funding alone. This is one area in which personnel with funding specifically assigned to these tasks is vital. A designated outreach person would interact with system developers to understand the function of the system well enough to explain it to stakeholders. In turn, they would interact with the user community not only to offer education about USGIN network resources, but also to get valuable feedback about USGIN products.

## Costs

In order to assess costs, the Working Group assembled an inventory of what components are likely to be necessary to sustain USGIN.

• Servers and software: (This list assumes that network infrastructure (internet connection, switches, firewalls, DNS, etc.) are in place)

o Catalog and repository server: Linux/Tomcat/PostgresQL/ GeoNetwork or Geoportal (could run on Windows stack as well). Only one is essential, any number is possible.

o To support a catalog system: a server for registries for identifiers and vocabularies, in addition to repositories for system specification documents and resources, like XML schemas that must be web accessible to support service operations.

o Data server: Windows/ArcGIS Server (dbms optional if shapefiles are used as data source), or Linux/Tomcat/PostGresQL/Geoserver or Mapserver. One server with modest capabilities could serve perhaps 300 Mb of data, depending on load.

• Personnel:

o Technical IT personnel: Need capabilities to deploy server software, load data, configure services, debug http traffic if there are problems. Data preparation (if standard interchange formats are being used) requires understanding of ETL using SQL queries and other techniques. Some understanding of XML and XML schema is required occasionally; GeoServer requires mapping from XML to database fields in an XML configuration. Someone with understanding of metadata content models and encoding is likely to be essential to get a catalog system working well.

o Outreach and marketing: personnel dedicated to the production and maintenance of documentation and educational materials, as well as face-to-face and online training programs.

• Maintenance:

o Individuals dedicated to system management, including arrangement of meetings, maintenance of hardware, user help lines, and network operation.

Software development and testing:

o Components for service deployment may be necessary if off the shelf solutions do not meet all requirements; applications for service and conformance testing, development of metadata, and network monitoring

Transition from current stovepipe data discovery and delivery systems to a loosely coupled, service based architecture can probably be done by large organizations like the USGS with the personnel and hardware they currently utilize, following whatever IT/hardware refresh cycles they currently use. Additions to hardware capacity in the form of servers, bandwidth, and online storage will be necessary to bring new data online, but this would be true no matter what approach to data delivery is adopted. The most significant additional investment will likely be in education of personnel who develop and deploy data and metadata services, in addition to human effort for data integration, documentation, and migration of existing data to new formats and delivery protocols. These investments will provide long-term return in staff capabilities if, as anticipated, these become widely accepted standard operating procedures. An additional return is the expected increase in utilization of the information resources and greater visibility for the agency providing data.

## Summary

The future of USGIN depends on taking measured steps to build this initiative and see it succeed. A summary of recommended actions from this document include:

Develop and deploy a marketing strategy to increase engagement of the State Surveys and stakeholders and cement a strong user base supporting USGIN.

Establish a governance structure led by a technical working team with representatives from USGS, State Geological Surveys, and other geosciences-focused communities

Decide if USGIN should be established as an independent non-profit 501(c)(3) foundation or a subdivision of an existing organization like OGC or ESIP

Conduct a comparative analysis of preferred sustainability models and select one to implement that will provide a conduit for financial resources to support infrastructure, a 2-3 person staff, and servers for system infrastructure

Inventory the State and Federal Geological Survey information architecture landscape to identify and prioritize use cases, and to identify capabilities and expertise to develop prototypes and testbed opportunities.

Work to promote consistent best practices

Engage with ESIP, NGC, DataOne18 and other geosciences communities to promote interests of geological surveys in national cyber-infrastructure development.

Select test bed opportunities to engage the geological survey community in the context of CDI, ESIP, NGC, OGC, DataOne and evolving NSF EarthCube activities.

## References

Percival, George, Editor, 2006-03-20, Interoperability Testbed Policies and Procedures: Open Geospatial Consortium Inc., document OGC 05-129r1.

Shapiro, Carl, 2000, Setting Compatibility Standards: Cooperation or Collusion?, *in* Dreyfuss, M. C., Zimmerman,

D. L., and First, Henry, Expanding the boundaries of Intellectual Property: Oxford, UK, Oxford University Press, p. 81-102.

DIN, 2000, Economic benefits of standardization, Summary of results: DIN German Institute for Standardization, e.V., 39 pages, accessed at http://www.din.de/sixcms\_upload/media/2896/economic\_benefits\_standardization.pdf 2011-10-03.

Blind, Knut, Jungmittag, Andre, and Mangelsdorf, Axel, 2011, The economic benefits of standardization, DIN German Institute for Standardization, e.V., 24 p., accessed at http://www.din.de/sixcms\_upload/media/2896/GNN\_2011\_engl\_FINAL.111681.pdf 2011-10-03.

Web sites:

18 https://dataone.org

Overall Benefits of Financial Data Standardization: http://www.gs1us.org/sectors/financial\_services/overall\_benefits\_of\_financial\_data\_standardization (accessed 2011­10-03)

Paper Submitted by: Stephen Richard, Arizona Geological Survey

1. “What is CUAHSI?” Consortium of Universities for the Advancement of Hydrologic Science, Inc., accessed June 1, 2012, http://www.cuahsi.org/docs/What-is-CUAHSI.pdf. [↑](#footnote-ref-2)
2. Ibid. [↑](#footnote-ref-3)
3. “Governance,” Consortium of Universities for the Advancement of Hydrologic Science, Inc., Accessed, April 16, 2013, http://www.cuahsi.org/Governance.aspx [↑](#footnote-ref-4)
4. Incorporated Research Institutions for Seismology (IRIS), “About IRIS,” last modified 2012, <http://www.iris.edu/hq/about_iris>. [↑](#footnote-ref-5)
5. Ibid. [↑](#footnote-ref-6)
6. “Governance,” Incorporated Research Institutions for Seismology (IRIS), accessed April 16, 2013, <http://www.iris.edu/hq/about_iris/governance>, [↑](#footnote-ref-7)
7. IRIS, “Membership, accessed April 16, 2013, <http://www.iris.edu/hq/about_iris/membership>, [↑](#footnote-ref-8)
8. Mohan Ramamurthy (Unidata Director), personal communication, April 25, 2013. [↑](#footnote-ref-9)
9. Ibid. [↑](#footnote-ref-10)
10. “Governing Committees,” Unidata, accessed April 15, <http://www.unidata.ucar.edu/community/index.html#governance>, [↑](#footnote-ref-11)
11. Ibid. [↑](#footnote-ref-12)
12. <http://www.geosamples.org/>, accessed April 15, 2013 [↑](#footnote-ref-13)
13. <http://www.geosamples.org/principles>, accessed April 15, 2013 [↑](#footnote-ref-14)
14. Lehnert, Kerstin, personal communication, April 25, 2013. [↑](#footnote-ref-15)
15. <http://www.opentopography.org/index.php/about/>, accessed April 15, 2013 [↑](#footnote-ref-16)
16. “History,” Open Topography, <http://www.opentopography.org/index.php/about/history>, accessed April 17, 2013. [↑](#footnote-ref-17)
17. <http://www.opentopography.org/index.php/about/partners>, accessed April 15, 2013 [↑](#footnote-ref-18)
18. <http://www.opentopography.org/index.php/about/advisory>, accessed April 15, 2013 [↑](#footnote-ref-19)
19. OneGeology, “Objectives,” <http://www.onegeology.org/what_is/objective.html>, accessed May 22, 2013. [↑](#footnote-ref-20)
20. OneGeology, “How,” <http://www.onegeology.org/what_is/how.html>, accessed May 22, 2013. [↑](#footnote-ref-21)
21. OneGeology, “Brief technical overview for the non-specialist,” <http://www.onegeology.org/technical_progress/home.html>, accessed May 22, 2013. [↑](#footnote-ref-22)
22. OneGeology, “How,” <http://www.onegeology.org/what_is/how.html>, accessed May 22, 2013. [↑](#footnote-ref-23)
23. OneGeology, “Organisation,” <http://www.onegeology.org/organisation/home.html>, accessed May 22, 2013. [↑](#footnote-ref-24)
24. “Foundation,” Apache Software Foundation, accessed April 16, 2013, <http://www.apache.org/foundation/>. [↑](#footnote-ref-25)
25. “How it Works,” Apache Software Foundation, accessed April 16, 2013, <http://www.apache.org/foundation/how-it-works.html>. [↑](#footnote-ref-26)
26. “How it Works,” Apache Software Foundation, <http://www.apache.org/foundation/how-it-works.html>, accessed April 16, 2013 [↑](#footnote-ref-27)
27. “Sponsorship,” Apache Software Foundation, <http://www.apache.org/foundation/sponsorship.html>, Accessed April 16, 2013 [↑](#footnote-ref-28)
28. DataONE, “What is DataONE?” Accessed April 17,2013, <http://www.dataone.org/what-is-dataone>. [↑](#footnote-ref-29)
29. <http://www.dataone.org/what-dataone>, accessed April 16, 2013. [↑](#footnote-ref-30)
30. Ibid. [↑](#footnote-ref-31)
31. <http://www.dataone.org/external-advisory-board>, Accessed April 17, 2013 [↑](#footnote-ref-32)
32. <http://www.dataone.org/what-dataone>, accessed April 17, 2013 [↑](#footnote-ref-33)
33. DataONE, “DataONE Organization,” accessed June 5, 2012, http://www.dataone.org/organization. [↑](#footnote-ref-34)
34. W3C Consortium, “About W3C,” last modified 2012, <http://www.w3.org/Consortium/>. [↑](#footnote-ref-35)
35. W3C Consortium, “W3C Mission,” last modified 2009, <http://www.w3.org/Consortium/mission>. [↑](#footnote-ref-36)
36. W3C, Consortium, “Current Members,” last modified May 31, 2012, http://www.w3.org/Consortium/Member/List [↑](#footnote-ref-37)
37. W3C Consortium, Thank You, W3C Funders, http://www.w3.org/Consortium/nmfunds [↑](#footnote-ref-38)
38. W3C Consortium, “Facts About W3C,” last modified 2012, http://www.w3.org/Consortium/facts. [↑](#footnote-ref-39)
39. Ibid. [↑](#footnote-ref-40)
40. W3C Consortium, “Facts About W3C,” last modified 2012, http://www.w3.org/Consortium/facts. [↑](#footnote-ref-41)
41. Ibid. [↑](#footnote-ref-42)
42. http://www.dpn.org/about/ [↑](#footnote-ref-43)
43. OPeNDAP, “What is OPeNDAP?” <http://www.opendap.org/>, accessed May 22, 2013 [↑](#footnote-ref-44)
44. OPeNDAP, “About”, <http://www.opendap.org/about>, accessed May 22, 2013. [↑](#footnote-ref-45)
45. Open Knowledge Foundation, “About Us,” <http://okfn.org/about/>, accessed May 22, 2013. [↑](#footnote-ref-46)
46. Ibid. [↑](#footnote-ref-47)
47. Open Knowledge Foundation, “Community,” <http://okfn.org/about/>, accessed May 22, 2013. [↑](#footnote-ref-48)
48. Open Knowledge Foundation, “Open Knowledge Foundation Governance,” <http://okfn.org/OKFgovernance/>, accessed May 22, 2013. [↑](#footnote-ref-49)
49. Open Knowledge Foundation, “Organisational information,” <http://okfn.org/about/>, accessed May 22, 2013. [↑](#footnote-ref-50)
50. Open Knowledge Foundation, “Board of Directors,” <http://okfn.org/about/board/>, accessed May 22, 2013. [↑](#footnote-ref-51)
51. Open Knowledge Foundation, “Open Knowledge Foundation Governance.” [↑](#footnote-ref-52)
52. DataBasin, “Homepage,” <http://databasin.org/>, accessed May 22, 2013. [↑](#footnote-ref-53)
53. Ibid. [↑](#footnote-ref-54)
54. DataCite, “What do we do?” <http://www.datacite.org/whatdowedo>, accessed May 22, 2013. [↑](#footnote-ref-55)
55. DataCite, “What is DataCite?” <http://www.datacite.org/whatisdatacite>, accessed May 22, 2013. [↑](#footnote-ref-56)
56. DataCite, “Faqs,” <http://www.datacite.org/faqs>, accessed May 22, 2013. [↑](#footnote-ref-57)
57. DataCite, “How to become a DataCite Member,” <http://www.datacite.org/Membership>, accessed May 22, 2013. [↑](#footnote-ref-58)
58. DataCite, “How to become a DataCite Member,” <http://www.datacite.org/Membership>, accessed May 22, 2013. [↑](#footnote-ref-59)
59. Internet Engineering Task Force, “About the IETF,” <http://www.ietf.org/about/>, accessed May 23, 2013. [↑](#footnote-ref-60)
60. Internet Engineering Task Force, “Mission Statement for the IETF,” <http://www.ietf.org/rfc/rfc3935.txt>, accessed May 23, 2013. [↑](#footnote-ref-61)
61. Internet Engineering Task Force, “Mission Statement for the IETF,” <http://www.ietf.org/rfc/rfc3935.txt>, accessed May 23, 2013. [↑](#footnote-ref-62)
62. “Internet Engineeering Task Force,” Wikipedia, <http://en.wikipedia.org/wiki/Internet_Engineering_Task_Force>, accessed May 23, 2013. [↑](#footnote-ref-63)
63. “Internet Engineeering Task Force,” Wikipedia, <http://en.wikipedia.org/wiki/Internet_Engineering_Task_Force>, accessed May 23, 2013. [↑](#footnote-ref-64)
64. Internet Engineering Task Force, “IETF NomCom,” <https://www.ietf.org/nomcom/>, accessed May 23, 2013. [↑](#footnote-ref-65)
65. ESIP Federation, “History,” last modified August 6, 2011, <http://www.esipfed.org/history>. [↑](#footnote-ref-66)
66. Carol Meyer (ESIP Executive Director), personal communication, June 5, 2012. [↑](#footnote-ref-67)
67. ESIP Federation, “History,” last modified August 6, 2011, <http://www.esipfed.org/history>. [↑](#footnote-ref-68)
68. “About OGC,” Open Geospatial Consortium, accessed April 17, 2013, http://www.opengeospatial.org/ogc [↑](#footnote-ref-69)
69. “OGC History (abbreviated),” OGC, Carol Meyer, personal email, June 5, 2012. <http://www.opengeospatial.org/ogc/history>, accessed April 17, 2013 [↑](#footnote-ref-70)
70. David Arctur, personal email, June 6, 2012. [↑](#footnote-ref-71)
71. “Programs,” Open Geospatial Consortium, last modified 2012, http://www.opengeospatial.org/ogc/programs.html. [↑](#footnote-ref-72)
72. Ibid. [↑](#footnote-ref-73)
73. “Policies,” Open Geospatial Consortium, last modified 2012, http://www.opengeospatial.org/ogc/policies/. [↑](#footnote-ref-74)
74. David Arctur, personal email, June 6, 2012. [↑](#footnote-ref-75)